DETECTING INFORMAL SETTLEMENTS FROM HIGH RESOLUTION IMAGERY USING AN OBJECT-BASED IMAGE APPROACH

by
KHALEED BALLIM

Thesis presented in fulfilment of the requirements for the degree of Master of Science in Geoinformatics at Stellenbosch University.

Supervisor: MR. NK POONA

December 2016

DECLARATION

By submitting this report electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 5 January 2016

ABSTRACT

The aim of this study was twofold: evaluate different approaches to deriving normalised digital surface models (nDSM), and develop a robust and transferable methodology for mapping informal dwellings. In the first component, three approaches to extract nDSMs were investigated: (i) light detection and ranging (LiDAR) data, (ii) high resolution aerial photographs in a process of image matching, and (iii) a series of aerial images captured using a hand-held camera using structure from motion (SfM) techniques. SfM is a relatively new technique that has not been widely used for nDSM extraction. This study represented a first attempt at evaluating the three approaches, particularly for mapping informal dwellings. The accuracy of the respective nDSMs was evaluated using vertical profiles, area-based, as well as positional-based accuracy assessment metrics. This provided a clear indication of the robustness of each of the models. Results showed that an nDSM can be successfully extracted in an informal settlement for informal dwelling mapping. Overall LiDAR achieved the highest accuracy in all three accuracy assessments, showing its ability to handle the undefined and complex morphology of informal settlements. To further test the robustness of the nDSMs, each model was applied to an independent test site with varying dwelling density and achieved improved accuracies.

In the second component, the utility of high resolution WorldView-2 imagery and object-based image analysis (OBIA) techniques to develop a robust and transferable methodology for mapping individual informal dwellings in the City of Cape Town was tested. A systematic approach was used to objectively identify segmentation and classification parameters. The supervised segmentation parameter tuner (SPT) tool was used to derive optimal segmentation parameters, and was evaluated using an area-based accuracy assessment which resulted in high compactness (> 86%) and correctness (>88%). To reduce data dimensionality and optimize the classification process, the RF algorithm reduced the original WV-2 feature set (n=40) and aerial imagery (n=60) feature sets by 23% and 53%, whereas the CART algorithm reduced the same feature set by 95% and 91% respectively. For classification, a supervised approach was adopted using the random forest (RF) algorithm, and a rule-based classification using a rule set in eCognition software. Although different feature subsets were selected by the RF and CART algorithm for the WV-2 and aerial imagery, similar classification accuracies were achieved in all the test sites.

KEY WORDS

Structure from motion, image matching, LiDAR, object-based image analysis, Boruta, CART, random forest, WorldView-2, informal settlements

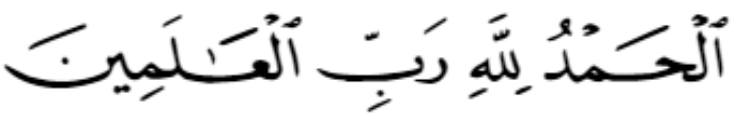
OPSOMMING

Die doel van hierdie studie was tweeledig: om die verskillende benaderings tot die skepping van genormaliseerde digitale oppervlak modelle (nDOM) te evalueer en om 'n robuuste en oordraagbare metodologie te ontwikkel om informele nedersettings te karteer. In die eerste komponent is drie benaderings tot die onttrekking van 'n nDOM ondersoek: (i) "light detection and ranging" (LiDAR) data, (ii) hoë resolusie lugfoto's deur gebruik te maak van 'n beeld-bypassingsproses en (iii) 'n reeks lugfoto's met 'n handgreep kamera geneem van "structure from motion" (SfM) tegnieke gebruik te maak. "Structure from motion" is 'n nuwe tegniek wat nog nie oor die algemeen gebruik word om nDOM te verkry nie. Hierdie studie is 'n eerste poging om die drie benaderings te evalueer, met die spesifieke doel om informele nedersettings te karteer. Die akkuraatheid van die onderskeie nDOM is geëvalueer van vertikale profiele, area-gebaseerde sowel as posisioneel-gebaseerde akkuraatheidsassessering statistieke. Dit het die robuustheid van elk van die drie modelle duidelik uitgewys. Die resultate dat 'n nDOM van 'n informele nedersetting met sukses verkry kan word en kan sodoende hierdie gebiede karteer. LiDAR het algeheel die hoogste akkuraatheid behaal tydens al drie akkuraatheidsevaluasies, wat hierdie metode se vermoë om die ongedefinieerde en komplekse morfologie van informele nedersettings te hanteer uit wys. Elke model is op 'n onafhanklike toetsgebied met wisselende woningsdigthede toegepas om die robuustheid van elke nDOM verder uit te lig. In die tweede komponent van die studie word die bruikbaarheid van hoë-resolusie WorldView-2 beelde en objek-gebaseerde beeld analise (OBIA) tegnieke om 'n robuuste en oordraagbare metodologie om individuele informele nedersettings in Stad Kaapstad te karteer getoets. 'n Sistematiese benadering is gebruik om segmentasie en klassifikasie parameters te identifiseer. Die gerigte "segmentation parameter tuner" (SPT) is gebruik om optimalesegmentasie parameters te verkry en die akkuraatheid van die segmentasie is geëvalueer deur gebruik te maak van 'n area-gebaseerde akkuuraatheidsassessering wat gelei het to hoë kompaktheid (> 86%) en korrektheid (>88%). Om data dimensionaliteit te verminder en die klassifikasie proses te optimeer, is die RF algoritme gebruik om die oorspronklike WorldView-2 voorwerp kenmerkstelle (n=140) en oorspronklike aerial voorwerp kenmerkstelle (n=60) ekwivalent aan 'n dimensionaliteitsvermindering van onderskeidelik 23% en 53%, terywl die CART algoritme 'n dimensionaliteitsvermindering van onderskeidelik 95% end 91%. 'n Gerigte benadering tot klassifikasie is gevolg deur gebruik te maak van die "random forest" (RF) algoritme, asook 'n reël-gebaseerde klassifikasie in eCognition sagteware. Om die robuustheid en oordraagbaarheid van die modelle te assesseer, is elke model op twee onafhanklike gebiede getoets.

TREFWOORDE

Structure from motion, beeldbypassing, LiDAR, objek-gebaseerde beeldanalise, Boruta, CART, random forest, WorldView-2, informele nedersettings

ACKNOWLEDGEMENTS



All praise and thanks to Almighty God.

I am very thankful for my parents and family for their support.

My heartfelt thanks to my friends on campus who never failed to say a motivational word in trying times.

I would like to convey my acknowledgements to the staff of the Department of Geography and Environmental Studies for constructive criticism during feedback sessions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

CART	Classification and regression trees
CGA	Centre for geographical analysis
CORC	Community organisation resource center
CS	Community survey
CSIR	Council for scientific and industrial research
DEM	Digital elevation model
DR	Discrete return LiDAR
DSM	Digital surface model
DTM	Digital terrain model
FEDUP	Federation of the urban poor
FW	Full-waveform LIDAR
GA	Genetic algorithm
GCP	Ground control points
GCP	Ground control point
GEOBIA	Geographic object-based image analysis
GHS	General household survey
GPS	Global positioning system
GSD	Ground sample distance
GSO	Global slum ontology
IES	Income and expenditure survey
IMU	Inertial measurement unit
ISDA	Informal settlement database atlas
ISN	Informal settlement network
ISUP	Informal settlement upgrading programme
LiDAR	Light detection and ranging

MGD	Millennium development goals
MRS	Multiresolution segmentation
MRS	Multiresolution segmentation
MVS	Multiview-stereo
NDH	National department of human settlements
NDHS	National department of human settlements
nDSM	normalised digital surface model
OBIA	Object-based image analysis
OOA	Object-oriented analysis
PSUP	Participatory slum upgrading programme
PUC-RIO	Electrical engineering department at the catholic university of Rio de Janeiro
RANSAC	Random sampling consensus
RANSAC	Random sampling consensus
RF	Random forest
SANSA	South African national space agency
SBC	Eskom spot building count
SDI	Slum dwellers international
SFM	Structure from motion
SIFT	Scale invariant feature transformation
SIFT	Scale invariant feature transformation
SPT	Segmentation parameter tuner
Stats SA	Statistics South Africa
TLS	Terrestrial laser scanner
UAVs	Unmanned aerial vehicles
VHR	Very high resolution

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND TO THIS STUDY

Rapid urbanisation and migration of people from rural to urban areas have led to urban expansion in the form of peripheralisation – the development of informal settlements in peri-urban areas (UN-Habitat 2012/2013). Those migrating to urban centres move in pursuit of socio-economic opportunities and improved livelihoods. However, developing economies like South Africa lack the necessary infrastructure to meet these demands, which results in migrants finding themselves unemployed and marginalised from both access to basic services and housing opportunities (Allen & Heese 2013). According to the UN-Habitat (2012/2013), informal settlements tend to be characterised by deplorable living and environmental conditions, overcrowded and dilapidated housing, and insecurity of tenure, which have resulted in the urban poor surviving in a condition of informality. Nonetheless, informal settlements are characterised by significant personal investment in dwellings, strong social structures, and effective community leadership (The Housing Development Agency 2012).

The terms used to describe informal settlements vary amongst countries and are used interchangeably (UN-Habitat 2015). The regional name in Brazil for informal settlements are *favelas*, *kampungs* in Indonesia, *chabolas* in Spain, *guetos* in Puerto Rico, *villas miseria* in Argentina, *hoovervilles* in the United States, *slums* in India and *informal settlements*, *shanty towns*, or *squatter camps* in South Africa (UN-Habitat 2015). The World Health Organisation (2013) stated that around 828 million people live in informal settlements, approximately one-third of the world's urban population. Likewise, the UN-Habitat (2012/2013) estimates that 863 million people worldwide are living in informal settlements, in contrast to 760 million in 2000 and 650 million in 1990. Sub-Saharan Africa is reported to have the fastest growing urban and informal settlement population, i.e. 4.53% and 4.58% respectively (UN-Habitat 2006).

The challenges of addressing informal settlements have been met with several international initiatives, including Target 11 of the Millennium Development Goals (MGDs) (Alliance 2015), to improve the lives of 100 million informal settlement dwellers by 2020. The Target 11 initiative has led to many informal settlement upgrading strategies and programmes in different countries. These strategies adhere to the Vancouver declaration on human settlements (1976), the Istanbul declaration on cities and other human settlements (1996), and the habitat agenda (1996).

The programmes include the ‘community urban development project’ and ‘community-based poverty reduction project’ in Nigeria, the ‘environmental improvement of urban slums program’ and ‘national slum development program’ in India, ‘participatory slum upgrading programme’ (PSUP) in African, Caribbean, and Pacific (ACP) group of states, and the ‘informal settlement atlas’, ‘upgrading informal settlement programme (UISP)’, and ‘re-blocking policy’ in South Africa.

A report compiled at the Expert Group Meeting on Slum Identification and Mapping (Sluizas et al. 2008) underlines the contextual conditions that can be found globally in informal settlements. The report states that within one city intra-informal settlement diversity can be found that requires methodological adjustments, and that it is necessary to understand both the nature of building construction characteristics and the development stage of informal settlements. Furthermore, it was concluded that there is currently no universal model or standard method for informal settlement identification and mapping, as the dynamic spatial-temporal behaviour, high inner-structural heterogeneity (Hoffman et al. 2008; Pinho et al. 2011; Shekhar 2012), and microstructure of informal settlements do not allow generically applicable solutions to be applied (Mpe & Orga 2014). The application of a generic solution to other informal settlements will result in data variations, and can cause large error margins in estimates that are disruptive in the evaluation of performance and intervention-based programmes (Kit 2013; Kohli et al. 2013).

1.2 PROBLEM FORMULATION

Obtaining up-to-date geospatial information for planning, identifying, and monitoring of informal settlements is required for the evaluation of performance and intervention-based programmes (Kohli et al. 2013; Kit et al. 2014). However, the inherent characteristics of informal settlements pose challenges for data collection, methodology development, and analysis (Hoffman et al. 2008; Pinho et al. 2011; Shekhar 2012).

Several attempts to map informal settlements at national, provincial, and at a city level have been conducted using surveys, census, and participatory-based approaches. The literature suggests that counting individual dwellings is the most reliable method (Kit & Lüdeke 2013); however, these approaches are inefficient as these are labour-intensive tasks, are limited by the accessibility of dwellings, and have inadequate quality assurance and monitoring processes, which result in under-coverage during enumeration.

Remote sensing-based approaches offer an alternative solution for methodology development that can provide predictable and consistent results for the identification of informal settlements using very high resolution (VHR) imagery (Pinho et al. 2011). Advancements in image-processing techniques provide a flexible and useful method to map informal settlements, specifically individual dwellings. One of the several approaches that have the potential to address the relatively complex and undefined urban morphology of informal settlements is object-based image analysis (OBIA) (Hofmann et al. 2008; Salehi et al. 2011; Shekhar 2012; Kit et al. 2013; Kohli et al. 2013).

Several segmentation, feature selection, and classification algorithms are available that have not been investigated in previous literature for the purpose of informal settlement identification. These algorithms provide an objective quantitative approach for the selection of optimised segmentation and classification parameters, and reduced feature subsets in which redundant and irrelevant features have been removed (Hapfelmeier & Ulm 2013). Traditionally, the selection of segmentation and classification parameters, as well as the feature subset was user-defined, error-prone, and the interpretation is biased by human subjectivity (Baatz & Schäpe 2000; Benz et al. 2004; Hofmann, Strobl & Blaschke 2008; Arvor et al. 2013).

In addition, very little is known about the value of normalised digital surface models (nDSMs) for informal dwelling mapping, and limited research has been performed on the use of structure

from motion (SfM) for extracting nDSMs. Previous studies have either assessed Image matching and LiDAR (Moreira et al. 2013), Image matching and SFM (Andrews et al. 2013), or LiDAR and SfM (Maiellaro, Zonno & Lavalle 2015). The inclusion of an nDSM for informal settlement mapping can assist to discriminate between bare soil and (oxidised) metallic roof structures (Pinho et al. 2011), and eliminate the need for multiclass dwelling classification schemas (Hoffman 2008; Ballim, Poona & Ismail 2014).

In summary, this thesis seeks to address five key questions:

1. Can a supervised segmentation approach be used to optimise segmentation parameters?
2. Can feature selection be used to determine which the most important features are?
3. Which classification algorithm will produce the highest accuracies?
4. Does the inclusion of WorldView-2 imagery increase the classification accuracies?
5. Can LiDAR, Image Matching, and SfM nDSMs handle the undefined and complex morphology of informal settlements, and provide contextual information to assist dwelling classification?

1.3 AIM AND OBJECTIVES

The aim of this research is to develop a methodology for identifying and mapping informal dwellings within informal settlements.
The aim is divided into two components; addressed in the following objectives:

1. Assess the accuracy of LiDAR, SfM, and Image matching for extracting normalised difference surface models (nDSMs), which can be used for informal dwelling extraction; and

2. Develop a robust and transferable methodology to identify and map informal dwellings using high resolution satellite imagery within an object-based image analysis environment.

1.4 METHODOLOGY AND RESEARCH DESIGN

An overview of the research design is provided in Figure 1.1, with the specific steps taken in data analysis shown in Figure 1.2. The research comprises five phases, namely: 1) knowledge building, 2) planning, 3) execution, 4) evaluation, and 5) synthesis. An empirical research approach using primary quantitative data based on the two main objectives was followed.

The research approach adopted for timely completion of this research included undertaking an intensive knowledge-building phase, where an in-depth literature review of methodological approaches, concepts, and characteristics of informal settlements and dwellings was undertaken.

The next phase was the planning phase, where a conceptual framework and the research problem were developed. The key research questions, the specific techniques of analysis, and the required data and software needed were identified. Once the conceptual framework was completed, the methodological framework needed to complete the research aim in the execution phase, was prepared.

In the context of this study, Chapter 3 and 4 are structured as journal articles with stand-alone entities from the overall thesis, each chapter with its own literature review, aim and objectives, methods, results, and conclusion. The execution phase entails the data-processing stage, where the data needed were processed and the best model from Objective 1 was selected based on accuracy assessments that were used as an ancillary dataset for Objective 2 as shown in Figure 1.2.

In order to validate the findings of the research, the evaluation phase was undertaken, which involved the analysis of the results and the methodological approach taken, as discussed in Chapter 3 and 4. Finally, in the synthesis phase, the research was critically assessed regarding the achievement of its stated objectives. The prospects and limitations of identifying informal dwellings using remote sensing and advanced image processing methods within informal settlements were discussed. The dissertation concludes with recommendations for further research.

A detailed explanation of the methods used is provided in Chapters 3 and 4.

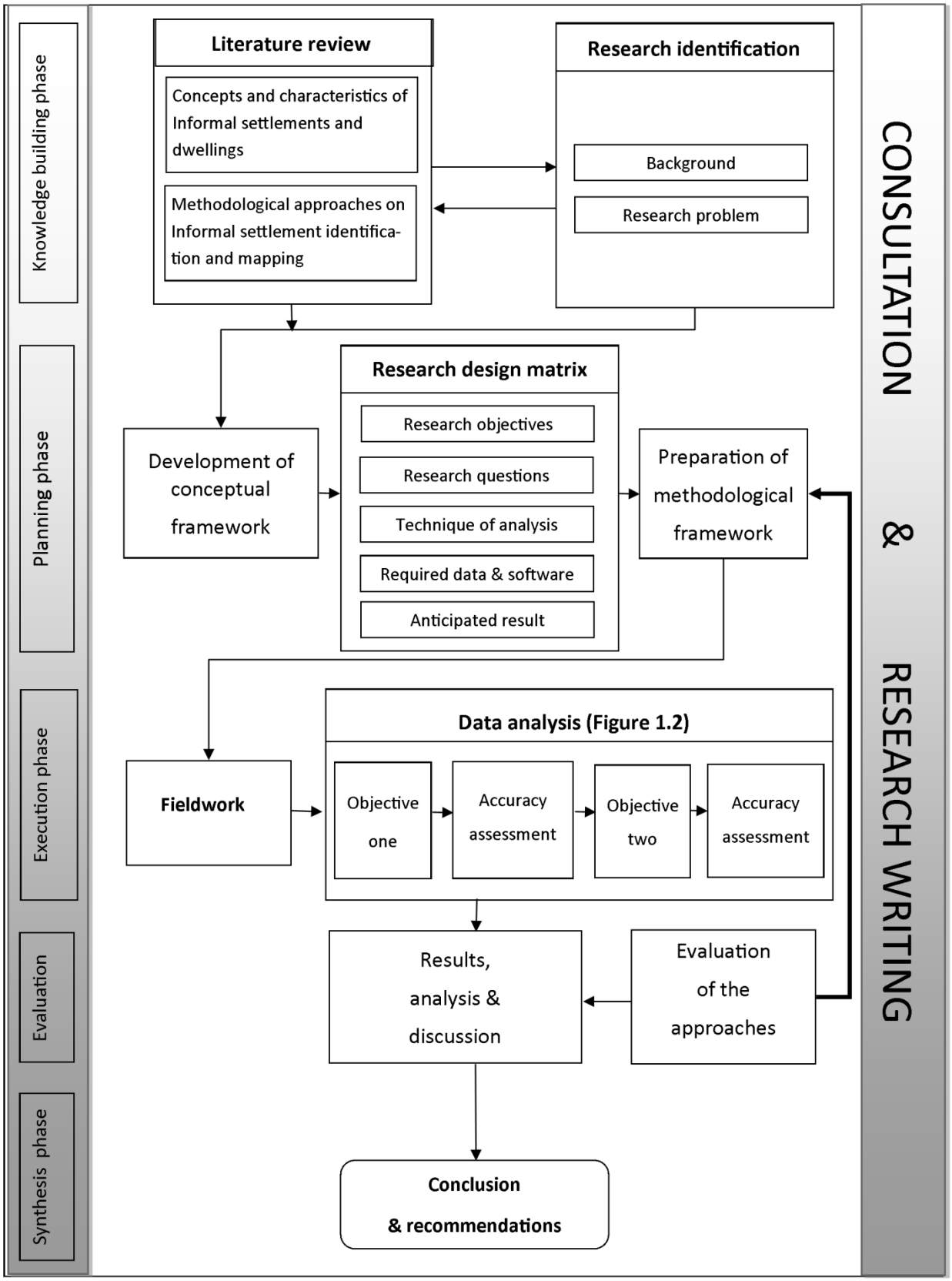


Figure 1.1 Research design diagram.

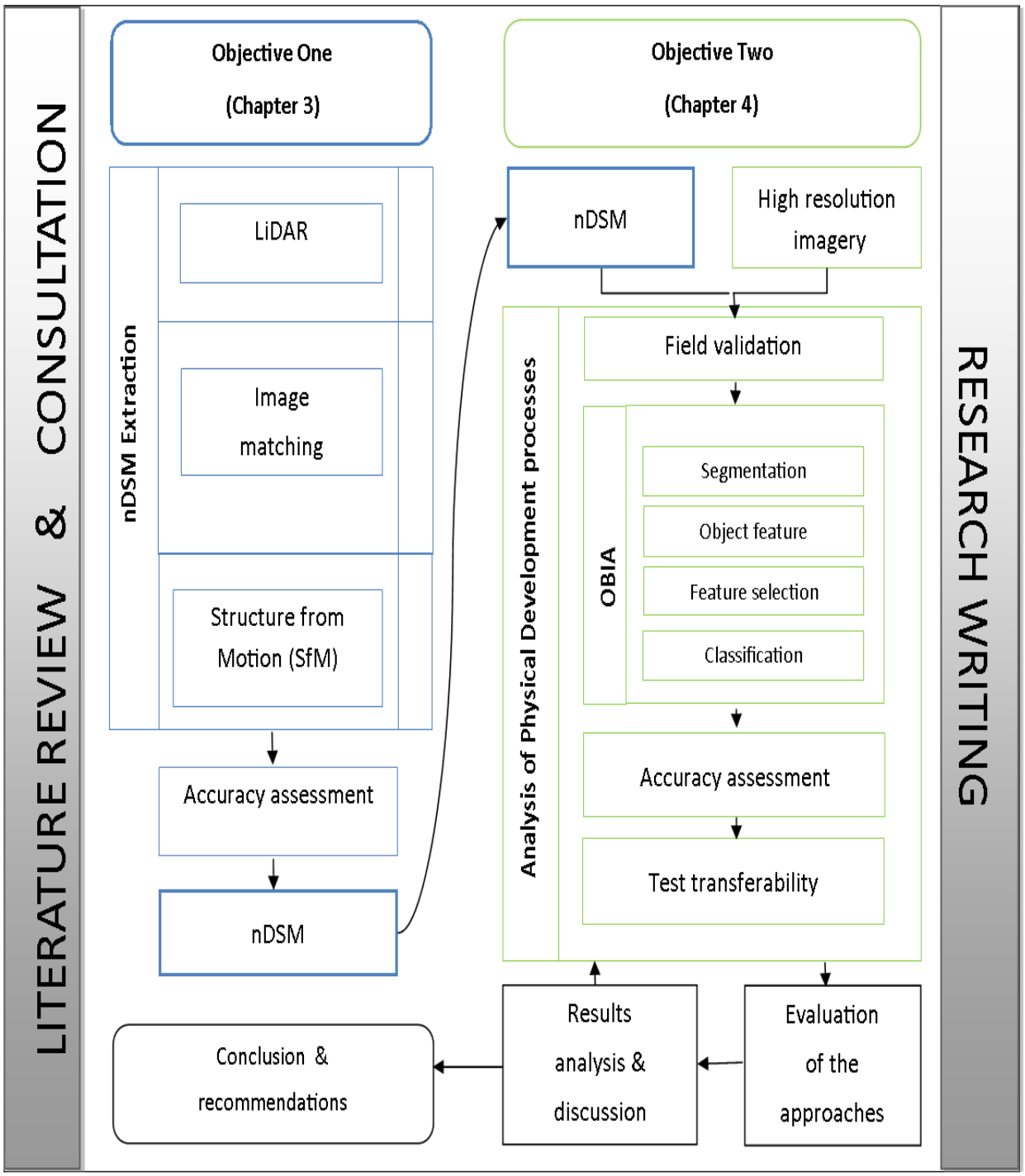


Figure 1.2 The research process provides more details with regard to how Objectives 1 and 2 will be carried out. The best nDSM generated from Objective 1 will be used as an ancillary dataset in Objective 2.

1.5 STUDY AREA

The study area, seen in Figure 1.3, is situated in the Western Cape, South Africa. Three different types of unstructured informal settlements were chosen for this study (see Figure 1.4). Structured dwellings are defined by occupying assigned plots of land with municipal facilities such as water, electricity, and waste removal provided, whereas unstructured dwellings, which are the focus of this study, are by definition those dwellings occupying any available land with minimal to no municipal facilities provided (Stasolla & Gamba 2007).

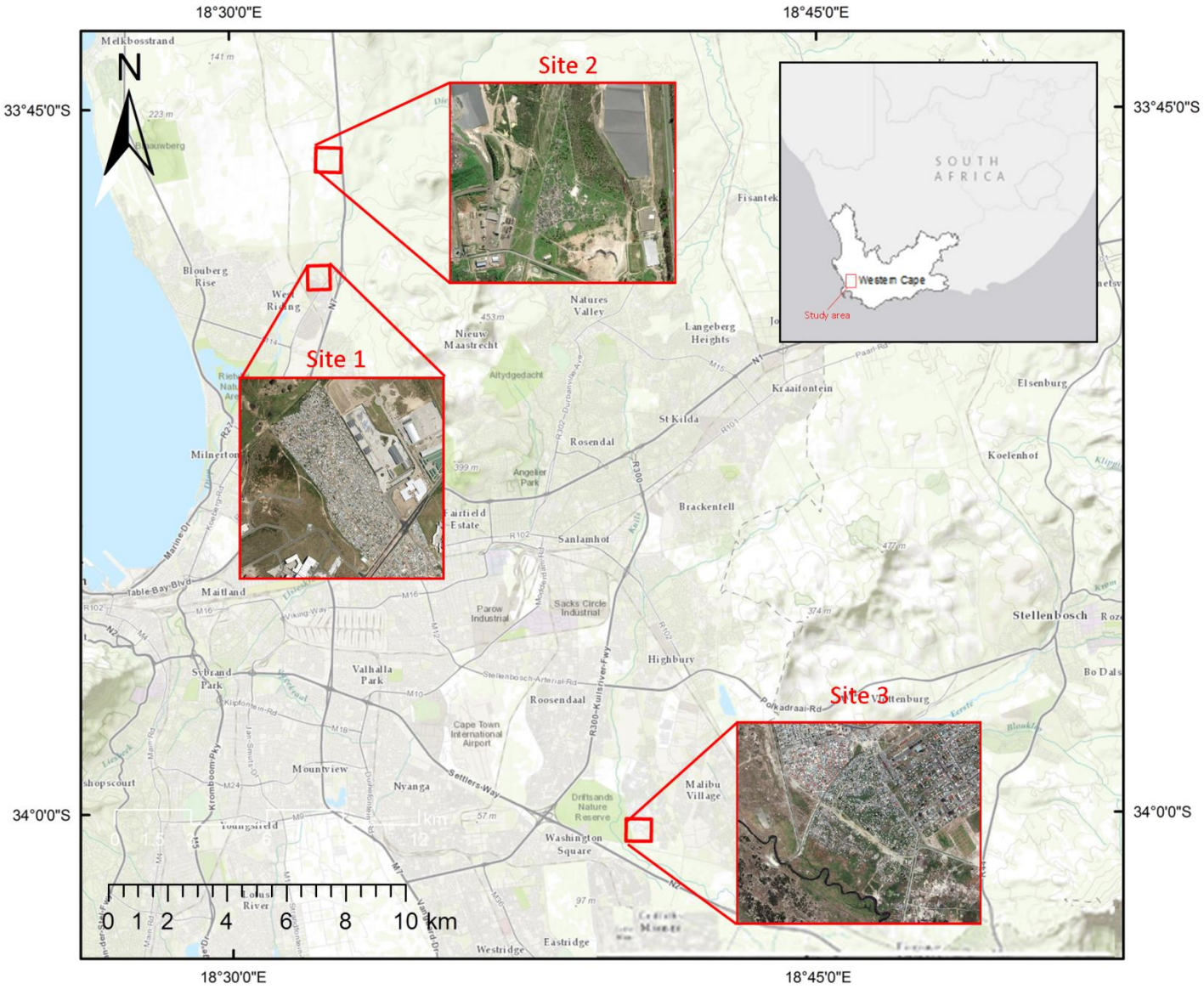


Figure 1.3 Location of study area with WorldView-2 inset overlaid on ESRI ArcMap imagery base map.

According to the Western Cape Department of Environmental Affairs and Development Planning (2009), building unit per hectare (bu/ha) can be expressed in South Africa as low density (n < 20), medium density (n < 30) or high density (30 < n < 50). Test site 1 has high density, with very crowded and clustered dwellings. Test site 2 is a relatively open area with

sparse dwellings and low density. Test site 3 is highly populated, in an open area, with medium density. The topography of all three study sites is fairly levelled. Study sites 1, 2, and 3 are consistent with properties described by Mason and Baltsavias (1997). The majority of the dwellings are single-storey structures, display simple geometry (four-sided), constructed from diverse materials with variable texture and colours, and with separation distance of ~1 m between dwellings.

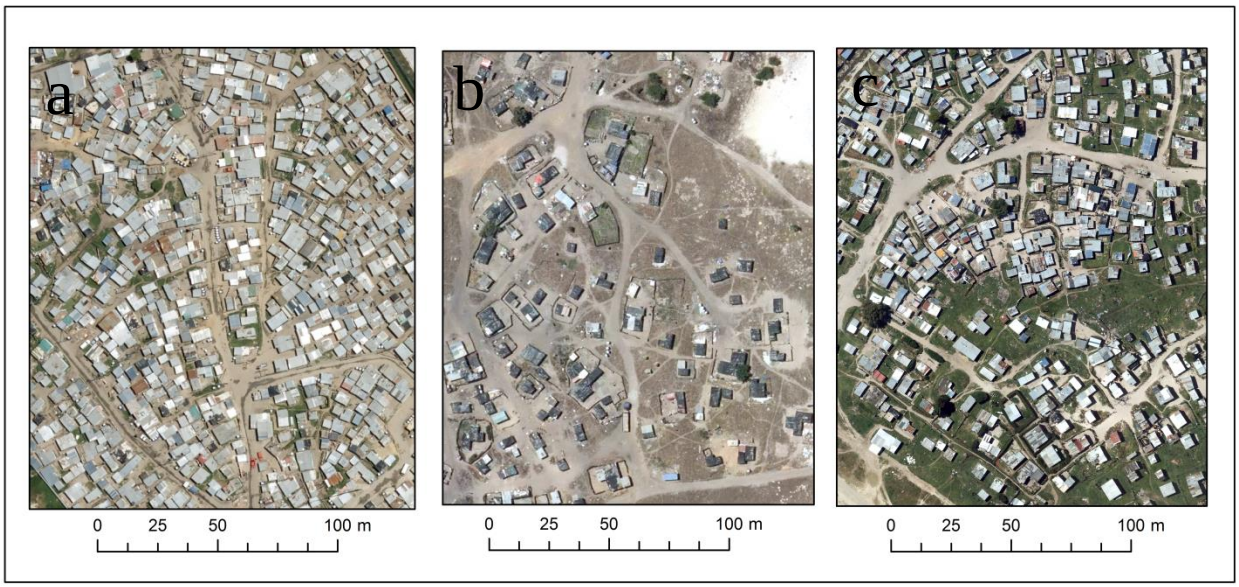


Figure 1.4 Study site 1–high density dwellings (a), study site 2–low density dwellings (b), and study site 3–medium density dwellings (c).

1.6 STRUCTURE OF THE THESIS

This thesis is presented in five chapters and is structured as follows:

Chapter 2 provides a review of the literature and defines important concepts for the study. Chapter 3 evaluates different approaches to extract nDSMs using LiDAR, Structure from motion (SfM), and image matching (IM) for informal settlement mapping.

Chapter 4 assesses the value of high resolution imagery, feature selection, and an object-based image analysis (OBIA) approach to develop a robust and transferable methodology for mapping individual informal dwellings.

Chapter 5 evaluates the overall results of the research, draws conclusions, and makes recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Approaches to mapping informal settlements

The reliable identification and monitoring of informal settlements have always been a difficult task for urban administrators in the developing world (Kit 2013). Several attempts to map informal settlements at national, provincial, and city level have been conducted. However, the lack of alignment regarding data collection methodologies and census inconsistencies, as discussed below, result in large error margins (Kit 2013; The Housing Development Agency 2012). Currently-used informal settlement enumeration approaches include (i) the census-based approach, (ii) participatory methods, and (iii) advanced remote sensing and image analysis-based methods (Kohli et al. 2012).

2.2 Survey and census-based approach

Various surveys and censuses are conducted by Statistics South Africa (Stats SA) including the community survey (CS), the general household survey (GHS), the income and expenditure survey (IES), and the national census. Surveys and censuses provide vital data for informed decision making in terms of policy formulation, programming, implementation and evaluation of projects (Census 2011). For the above to be achieved, the data have to be accurate and relevant. However, one of the major challenges is related to census under-coverage as a result of incomplete and inaccurate address lists and inadequate quality assurance and monitoring processes (Stats SA 2011). Furthermore, census under-coverage is associated with areas that lack infrastructure such as informal settlements. The absence of censuses in informal settlements in the Western Cape is largely due to the inaccessibility of make-shift structures due to political intolerance or general disorder. This makes application of census-listing methodology impractical, which results in certain structures being erroneously excluded during enumeration (Census Undercount and Strategies 2011). Additionally, a census is restricted by temporal collection gaps, the time needed to check raw data, and to collate, and present derived statistics to users (Kohli et al. 2013). Consequently, very few informal settlements have been included in censuses – which is an indication of the lack of effective detection and monitoring methods applied to include informal settlements.

The City of Cape Town estimated the number of dwellings in informal settlements based on the 2001 Census count of 110 000 as a baseline, with an estimated annual growth rate of 4.51%. As a result, an estimated 224 000 dwellings is predicted for 2015 (Housing Development Agency

2012). Despite the lack of census data regarding informal dwellings, as highlighted in the census undercount and strategies, the City of Cape Town continues to use estimated baselines and implied growth rates for policy formulation and informal settlement management strategies.

2.3 Participatory-based approach

The participatory approach is a methodology that involves the cooperation of informal settlement dwellers in order to generate spatial information, household count, and develop a socio-economic profile of informal settlements (Kohli et al. 2012). According to UN-Habitat (2014), this multi-stakeholder platform promotes necessary partnerships, governance arrangements, institutional structures and financing options that result in inclusive planning and sustainable outcomes. The participatory slum upgrading programme (PSUP), established by the UN-Habitat and the Slum Dwellers International (SDI), has championed the importance of community knowledge and encourages an inclusive environment, where communities can be empowered to become active partners with stakeholders in devising strategies to plan sustainable informal settlement upgrades (SDI 2013). The City of Cape Town supports the concept of 'active participation', 'dialogue', and 'continual engagement' with informal settlement communities.

In view of meaningful participation in South Africa, the Community Organisation Resource Center (CORC) was established (SDI 2013). The CORC provides support for urban and rural poor communities to mobilise themselves around their own resources and capacities. It advocates social processes by facilitating engagements with formal stakeholders such as the state and communities with regards to household upgrades and the provision of basic services. The CORC professionals provide contextually relevant learning and training that enable community-based planning and partnership for data collection and enumeration of informal settlements (South African SDI Alliance 2013). Subsequently, this has seen the launch of the Informal Settlement Network (ISN) which was developed from broad community experiences in developing a viable upgrade approach to the informal settlement model (South African SDI Alliance 2013). The Federation of the Urban Poor (FEDUP) is a nationwide federation of informal settlement dwellers that addresses the acquisition and incremental upgrade of informal settlements. The CORC, ISN, and FEDUP play a crucial role in building human capacity and in the steering of local projects, including 're-blocking' policy.

Since November 2013, the City of Cape Town has adopted the 're-blocking' policy for dense informal settlements (South African SDI Alliance 2013). Re-blocking refers to the reconfiguration and repositioning of informal dwellings within an informal settlement (in situ

rearrangement) to better utilise space for planning of provisions, installation of local government services, and disaster risk mitigation (South African SDI Alliance 2013). The 're-blocking' policy mobilises local communities to enumerate dwellings within their own informal settlement, and has seen recent success in Cape Town and Johannesburg (SDI 2013).

Kit and Lüdeke (2013) agree that counting individual dwellings is the most reliable method for informal settlement estimation. Although very accurate, this method is extremely time-consuming and effort-intensive, and thus an alternative method is required to reduce and streamline the enumeration process (Kohli et al. 2012).

2.4 Remote sensing-based approach

A remote sensing-based approach uses image-processing techniques and VHR imagery to map and monitor the spatial behaviour of informal settlements, and offer a worthy alternative to field data collection (Kit 2013). The National Department of Human Settlements (NDHS) commissioned the development of two atlases, namely the Human Settlement Atlas that was compiled by the Council for Scientific and Industrial Research (CSIR) and the Informal Settlement Atlas compiled by AfriGIS. According to the Housing Development Agency (2012), the 2009/2010 atlases were created from the data available within municipalities. Informal settlement boundaries were identified and digitised from available aerial and satellite imagery from different years up to 2006. In a related project, the North West Department of Human Settlements created the Informal Settlement Upgrading Programme (ISUP) using SPOT and IKONOS imagery to create a multiyear database of informal settlement boundaries.

Eskom's SPOT Building Count (SBC) mapped structures in South Africa using image interpretation and manual digitisation of SPOT 5 and aerial imagery. The SBC mapped identifiable building structures by point, but where informal settlements were too dense to determine the number of individual structures, they were mapped by polygons representing informal settlement boundaries. As a result, there are 234 polygons categorised as dense informal settlements in the Western Cape. The above mentioned projects do not provide an indication of the number individual dwellings in informal settlements, but rather a collective number of informal settlements and the total area covered in square kilometres (Housing Development Agency 2012). The only South African study to date that provides individual dwelling estimates is the Informal Settlement Database Atlas (ISDA) (2012) developed by the South African National Space Agency (SANSA). The ISDA mapped 140 informal settlements in the North West province, with an estimated 77 600 individual dwellings. Similar to Eskom's SBC, SANSA

used a combination of field-work and high resolution aerial photography to identify individual structures in informal settlements through image interpretation. A report of the Expert Group Meeting on Slum Identification and mapping (Sluizas et al. 2008), concluded that remote sensing-based approaches and VHR imagery provide a flexible and useful method to identify and map informal settlements. One of the several approaches that has the potential to address the relatively complex and undefined urban morphology of informal settlements is object-based image analysis (OBIA) (Hofmann et al. 2008; Salehi et al. 2011; Shekhar 2012; Kit et al. 2013; Kohli et al. 2013).

2.5 An overview of object-based image analysis

OBIA also referred as object-oriented image analysis (OOA), or geographic object-based image analysis (GEOBIA) seeks to bridge broader principles of remote sensing, image analysis, and GIS concepts (Blaschke 2010). OBIA is a semi-automated robust and flexible image processing approach that attempts to overcome the limited spectral information of pixels by combining spectral, spatial, and textural information to form homogenous image objects through segmentation (Baatz & Schäpe 2000; Blaschke 2010). The resulting image objects are used as input for the subsequent classification tasks (Belgiu et al. 2013).

2.5.1 Segmentation

Segmentation is undertaken to approximate meaningful landscape entities by representing the inherent patterns and mutual relationships between image objects (Saha, Wells & Munro-Stasiuk 2011; Dronova 2015). The aim of segmentation is to ensure the local homogeneity within image objects, while still representing the global heterogeneity within the image (Su et al. 2008). It is well understood that the accuracy and reliability of classification largely depends on the accuracy of the segmentation method and strategy (Baatz & Schäpe, 2000; Benz et al. 2004). A successful segmentation process should follow the following criteria: the sum of all individual objects must be equal to the whole image and objects must be mutually exclusive. For example, objects should not overlap and pixels in the same class should have similar values, and hence different classes should have dissimilar values (Janak 2010).

Mathematically, a segmentation procedure can be represented as:

$$I = R_1 \cup R_2 \cup \ldots \cup R_n \qquad \text{Equation 1.1}$$

Where I is the whole image;

$R_1, R_2 \ldots R_n$ represents non-overlapping contiguous individual regions; and

$\cup$ is the mathematical union.

Under-segmentation can occur if image objects contain low interior homogeneity and results in an image that contains multiple different semantic objects (Liu & Xia 2010), whereas over-segmentation occurs if image objects contain high interior homogeneity and low mutual heterogeneity (Zhang 2012). Both under- and over-segmentation can negatively affect classification as image objects do not represent landscape entities.

There are two basic segmentation principles: top-down segmentation that partitions the image into smaller objects, and bottom-up segmentation that merges smaller objects into bigger objects (eCognition 2012). Several segmentation algorithms are available in eCognition, the first commercially available OBIA software (Blaschke 2010); for example the popular Multiresolution segmentation approach proposed by Baatz and Schäpe (2000).

Multiresolution segmentation (MRS) has proven to be one of the most successful image segmentation algorithms in the OBIA framework (Witharana & Civco 2013), as it yields the most homogenous and morphologically representative objects (Mashimbye, de Clercq & Van Niekerk 2013). The MRS algorithm is a global bottom-up segmentation based on a pairwise region merging technique that seeks to minimise the average heterogeneity and maximise the respective homogeneity of the objects created (eCognition 2012). This is achieved by merging pixels iteratively in pairs, until a threshold of homogeneity is not exceeded locally amongst the collection of pixels. The algorithm looks for the best-fitting neighbours for potential mergers. If the best-fitting neighbour is mutual, image objects are merged. However, if there is no homogenous agreement, or no mutual neighbour, the best candidate image object becomes the new seed and finds its best partner. The matching iteratively continues until no further merging is possible (eCognition 2012).

Best-fitting neighbours are found based on the homogeneity criteria of scale, colour, and shape. Scale is regarded to be of greater significance than shape and colour (Pinho et al. 2012). Scale directly impacts the size of image objects and defines the maximum standard deviation of the homogeneity of the image objects (Su et al. 2008). The shape criterion is defined by the textural homogeneity of the image objects and is constituted by weighting compactness versus smoothness. Smoothness optimises how smooth image objects' boundaries are, whereas compactness optimises the overall compactness of the image objects (eCognition 2012). Modifying the value of the shape criteria it optimises spatial homogeneity (eCognition 2012). The colour criteria are the digital values of the image objects and represented as colour = 1-shape. The composition of the homogeneity criteria (shape and colour) is a weighted percentage

equalised to 1. Although MRS generates the most morphologically representative objects, it can be computationally intensive and unsuitable for larger datasets (Li et al. 2014).

Other segmentation algorithms available include chessboard segmentation, quadtree-based segmentation, and contrast split segmentation. Chessboard segmentation segments an image into square image objects but does not, however, produce representative objects as spectral values are not taken into account (Li et al. 2014). Similarly, quadtree segmentation represents an image as square image objects but of varying sizes. Contrast split segmentation segments an image into dark and bright image objects based on contrast. It is important to note that the time spent finding optimal segmentation homogeneity criteria, especially scale, impedes OBIA operational frameworks (Duro et al. 2012) as *scale* in eCognition is unitless and difficult to relate to spatial relationships (Hay et al. 2005).

Finding optimal homogeneity criteria for segmentation is time-consuming, subjective, and dependent on the analyst's experience (Salehi et al. 2012). Thus the segmentation process is considered to be a "black art" and in order to find a set of optimal segmentation parameters, the users have to adopt a trial-and-error approach until reasonable segmentation parameters are found or until the user does not want to continue testing using a trial-and-error approach (Zhang et al. 2010).

Several segmentation approaches have been developed and tested to objectively select optimal segmentation homogeneity criteria, such as by Baatz and Schäpe (2000), Feitosa et al. (2008), Zhang et al. (2008), Martha et al. (2011), Johnson and Xie (2011), and Drăguţ and Eisank (2011). For example, Drăguţ et al. (2010) created a generic automated segmentation tool that detects patterns in data called estimation of scale parameters (ESP). ESP automatically identifies patterns in data at three different scales, from finer to larger objects in a data-driven approach (Belgiu et al 2011). Zhang et al. (2010) developed the fuzzy-based segmentation parameter optimiser (fbSP). The fbSP is a supervised software tool that determines optimal segmentation parameters using fuzzy logic analysis. The segmentation parameters tuner (SPT) was developed at the Computer Vision Lab (LVC) of the Electrical Engineering Department at the Catholic University of Rio de Janeiro (PUC-RIO), and uses a supervised approach to optimise the Baatz segmentation parameters proposed by Baatz and Shäpe (2000) using a Genetic Algorithm (GA). It was found that the accuracy and reliability of the classification within OBIA is dependent on the image segmentation method and strategy (Belgiu et al.2013).

2.6 Feature selection

The number of available object features for classification makes a detailed qualitative exploratory analysis of every individual image feature extremely time-and-effort intensive, which consequently has led to the introduction of feature selection methods (Novack et al. 2008; Laliberte, Browning & Rango 2012). The addition of redundant and unnecessary features can lead to poor representation of real-world phenomena and the deterioration of classification accuracy known as Hughes phenomena or the "curse dimensionality", which can be avoided by reducing the data dimensionality of the given feature set (Hughes 1968).

Feature selection methods are aimed at improving classification accuracy by selecting an optimal subset of features in which redundant features have been removed (Hapfelmeier & Ulm 2013). It is also important to identify significant features as a comprehensive feature extraction methodology is the precondition for successful work with image objects (Nussbauma et al. 2008). There are two classes of methods for feature selections, which depend on where the feature selection method is placed in relation to the classification algorithm (Jain & Zongker 1997). Filter methods (Pudil, Novovicova & Kittler 1994) select a subset of features independent of the learning algorithm by eliminating irrelevant features by investigating the underlying distributions, whereas wrapper methods (Kohavi & John 1997) apply a learning algorithm in order to search for an optimal or near optimal subset of features. Feature selection algorithms can be assessed based on classification accuracies, ability to rank and reduce features and the ease of use (Laliberte, Browning & Rango 2012). A selection of the most common feature selection algorithms are presented below.

2.6.1 Classification tree analysis

Classification tree analysis (CTA) is a non-parametric classification approach to ranking features. A decision tree (De'ath & Fabricius 1999) provides a hierarchical representation of the feature space in which features are allocated to classes based on observations (De'ath & Fabricius 1999). CTA handles categorical and continuous data equally well and is most useful for data that have non-normal distributions. Friedman (2001) note that CTA can be adversely affected by complex datasets, inaccurate training data, and that outliers can potentially account for a large portion of variability in the data resulting in over-fitting. In addition the presence of an unbalanced data-set, with some classes more heavily represented than others, can affect the performance of the CTA. CTA has been shown to be an effective feature selection method and has been applied successfully in an OBIA environment (Chubey, Franklin & Wulder 2006; Laliberte et al. 2007; Addink et al. 2010; Laliberte, Browning & Rango 2012).

2.6.2 Feature space optimisation

Feature space optimisation (FSO) calculates optimal feature combinations based on training class samples (Laliberte, Browning & Rango 2010). FSO uses Euclidean distance to determine the best combination of object features in the feature space. FSO evaluates class separation distance – the largest of the minimum distances between the least separable classes (Leduc 2004; Aminipouri et al. 2012). The selected features are ranked based on the order of importance within the FSO tool. It is, however, considered a 'black box' approach as unclear feature ranking is given without defined rules. FSO can be computationally intensive with the inclusion of textures over and above spatial and spectral features, as optimisation is based on average distance, which may be globally small but can be locally large between classes (Laliberte, Browning & Rango 2010).

2.6.3 Random forest algorithm

The Boruta algorithm (Kursa & Rudnicki 2010; Kursa 2012) is a wrapper approach embedded with the Random forest classifier that gives a numerical estimate of the feature importance. Unlike other wrapper methods that find a minimal subset of features, Boruta selects both strongly and weakly relevant features (Kursa & Rudnicki 2011). Including strongly and weakly relevant features contributes to improved classification accuracy and provides the highest prediction accuracy. Kursa, Jankowski and Rudnicki (2010) presented an extensive review for Boruta as a system for feature selection.

2.7 Classification

Once the image has been segmented and the objects have been created, the success of a classification depends on several factors, including data, computational and operational requirements, the availability of training data, and the choice of a suitable classification procedure (Shang et al. 2009; Belgiu & Drăguţ 2016). Two traditional approaches to image classification include unsupervised and supervised classifiers.

Unsupervised classifiers cluster pixels into classes based on the identification of natural patterns within the feature set and do not require prior information or training (Campbell 2006). Unsupervised classifiers are relatively easy and fast to implement and perform best when information classes are spectrally distinct (Gao 2009). However it is not uncommon that spectral classes do not correspond to information classes, as naturally occurring clusters can drift away from class centres (Lee, Grunes & Pottier 2001), and as a result are used less than supervised classifiers.

Supervised classifiers are considered better than unsupervised approaches as these classifiers are able to learn the characteristics of target classes and incorporate prior knowledge from training samples. Samples used to train supervised classifiers need to fulfil the following: 1) training samples must be class balanced, 2) training samples must be representative of the target classes, and 3) training and validation must be statistically independent (Belgiu & Drăguţ 2016). However, since training samples are collected manually, certain problems may arise, introduced by human error such as small training sample sizes, which can cause the supervised classifier to lack discrimination and generalisation capabilities (Myburgh & Van Niekerk 2013).

Millard and Richardson (2015) note that a supervised classifier needs to be able to efficiently 1) mitigate the Hughes phenomenon, 2) deal with the nonlinearity of variables, 3) deal with imbalanced training samples, and 4) reduce computational time. Supervised machine learning algorithms have therefore become popular for their ability to train quickly and handle large datasets (Rodrigeuz-Galiano et al. 2012).

2.7.1 Decision tree classifier

Decision tree classifiers provide a hierarchical representation of a binary tree using a sample of training data. They are able to generate rules that can be easily understood and interpreted, ability to rank and reduce features, have the ability ability to handle nonlinear relationships between features, and classes and has low processing time and relatively high accuracy (Laliberte, Browning & Rango 2012). They output thresholds based on an entire classification tree by providing easy transfer to a rule-based classification (Laliberte et al. 2007). Decision tree classifiers have increased in popularity over traditional methods and has several advantages namely 1) they do not rely on any assumptions regarding the distribution of data, 2) are non-parametric, 3) a wide range of data sources can be used as inputs into classification, and 4) they handle continuous and categorical information equally well, whereas traditional classifiers cannot include categorical data (Lawrence et al. 2006).

Ghose, Pradhan and Ghose (2010) compared decision tree classifiers' land cover classifications of remote sensed satellite data with a traditional method namely as maximum likelihood classifier (MLC). The highest overall accuracy and kappa index were achieved by the decision tree classifiers (98% and 97%), whereas MLC achieved 95% and 94% respectively. Similarly, Pooja, Janyanth and Koliwad (2011) classified a multispectral satellite image (LISS III) using decision tree classifiers (87%), and evaluated its performance with MLC (82%). Decision tree classifiers' rules were simple to understand and implement and were less computationally intensive.

Decision trees were successfully implemented for feature selection by Laliberte, Browning & Rango (2012) who evaluated a decision tree classifier with the Jeffreys-Matusita distance (JM) and feature space optimisation (FSO) for object-based classification with digital aerial imagery. Decision tree classifiers were best suited for this particular type of imagery with numerous image classes because of the efficient workflow, easy interpretability, and the ability to both rank and significantly reduce features.

2.7.2 Random forest classifier

The Random Forest (RF) classifier (Breiman 2001) is an ensemble of weak unbiased classification or regression trees (Ismail & Mutanga 2010). Ensemble classifiers can be based on an individual supervised classifier or on a number of different supervised classifiers (Belgiu & Drăguţ 2016) that are trained using Bagging (Breiman 1996) or boosting approaches (Schapire, 1990; Freund & Schapire, 1997). RF is based on two techniques namely CART and Bagging. Bagging, also known as bootstrap aggregation, trains each classifier in the ensemble on a random subset of the training sample set, and has achieved greater accuracy than using a single classifier such as decision tree classifiers (Briem et al. 2002; Miao et al. 2012). Chan et al. (2012) noted that the RF classifier is best suited when a small sample size is used with high dimensional data inputs. RF requires two parameters to produce forest trees: 1) the number of decision trees to be generated (Ntree), and 2) the number of variables to be selected and tested for the best split when growing the trees (Mtry) (Belgiu & Drăguţ 2016). High variance and low bias (Breiman 2001) are ensured by growing the forest to the user-defined number of trees (Ntree). Mtry is usually set to the square root of the number of input variables, and Ntree is recommended to 500. However, Ghosh et al. (2014) set Mtry to the total number of available variables which resulted in a significant increase in computational time.

Breiman (2001) noted that the RF classifier has several advantages including:

1) it is relatively robust to outliers;
2) is has superior accuracy over other machine learning algorithms;
3) it gives useful internal estimates of error, strength, correlation, and variable importance;
4) it is computationally less intensive than other algorithms; and
5) It does not overfit because of the law of large numbers (Rodriguez-Galiano et al. 2012).

Du, Zhang and Zhang (2015) semantically classified buildings using the RF classifier. RF was capable of handling a large number of samples and high dimension and heterogeneous features. An overall accuracy of 71.50% and kappa index of 0.59 were achieved. Overall accuracy was greatly reduced by misclassification and uneven distribution of buildings in different categories.

Rodriguez-Galiano et al. (2012) evaluated the performance of the RF classifier and CTAs for a complex study area and various land cover classes. The results showed that the RF is a superior classifier as it allowed increased differentiation between different classes, achieving an overall of 92% compared to 86% achieved by CTA. Novack et al. (2011) found similar results, as RF produced the best overall accuracies (95%) amongst three classifiers such as regression trees (85%) and decision trees (77%), and SVM (57%), using WorldView-2 and Quickbird-2 simulated imagery in an object-based environment.

2.7.3 Rule-based approach

The rule-based classification or membership function classifier uses fuzzy or crisp membership functions and its logical operators to define membership of image objects (Myint et al. 2011). A rule-based classification can have a single or several conditions for assigning objects to a class (Dronova 2015) and is suited to handle vagueness and ambiguities in information extraction (Rahman & Saha 2008). A rule-based classification analyses the image objects in line with the set of formulated conditions (rule sets) to assign objects features to a class that best meets the defined specifications (Mathenge 2010). Rule-based classification relies on prior knowledge of the features of interest and allows the analyst to evaluate in detail the spectral similarities and differences between image objects (Xu 2013). However, building a rule set is a time-consuming task (Belgiu et al. 2014) as the number of image features greatly challenges the operators to determine the most relevant image features and thresholds.

In order to find relevant rulesets and corresponding image object features and thresholds, rule-based classification can intrinsically rely either on human knowledge (Myint et al. 2011; Kohli et al. 2012), by mimicking photo-interpreters knowledge (Sebari & He 2013), cognitive methods such as explicit rules from domain experts (Belgiu et al. 2014; Zhou et al. 2010), or feature selection methods.

Salehi et al. (2012) developed a hierarchical rule-based classification framework on a small subset of QuickBird imagery to classify a complex urban environment. An overall accuracy of 92% was achieved, and when applied to a larger subset of QuickBird and IKONOS imagery, 86% was achieved. The authors attribute the success of the rule-based classification to the use of expert knowledge in the development and selection of the relevant object features and thresholds.

Belgiu et al. (2014) evaluated the variability of rule-based classification carried out by three independent experts (referred to as C1-C3) on the same WorldView-2 satellite imagery. The overall results showed significant differences among all the experts as C1 achieved 87.3%, followed by C3 (80.7%), and C2 (78.24%). The performance of the developed rule set was tested

on a secondary site and a decrease in all classifications by C1 (82.29%), C2 (70.49%), and C3 (73.6%) was found. The difference in classification results between the experts was a result of different object features used, the definition of threshold intervals for the selected features and the allocation of hierarchical classification levels.

2.8 Classification of informal settlements

The few studies that have demonstrated the use of OBIA for classification of informal settlements have integrated the relative knowledge of real-world characteristics, such as the spectral, the geometric, and contextual properties, and relationships between objects (Hoffman 2001; Hoffman et al. 2008; Shekhar 2008; Kohli et al. 2012; Kohli et al. 2013).

A pioneering study by Hoffman (2001) used OBIA to identify informal settlements from IKONOS imagery in the City of Cape Town. Informal settlement classification was undertaken using sub-classes that described settlement forms (dense, medium, new and bright) based on complex hierarchy and class descriptions such as textural and spectral features .The author found that the ability to detect informal settlements was dependent on the spatial resolution of the imagery. No quantitative results were presented as the findings of the study.

This research was later improved by Hofmann et al. (2008), who showed that several modifications were required when applying extraction methods to a QuickBird scene in Brazil. The adaptions included simplified and pruned class-hierarchies to make the chosen class descriptors in theory more transferable to comparable scenes. The results of this study demonstrated that the selection of a strategy for informal settlement segmentation and classification is data and context-specific.

Tiede et al. (2010) extracted structures in refugee and IDP (internally displaced persons) camps in West Darfur using GeoEye-1 imagery. Classification was performed by incorporating dwelling spatial characteristics and limited use of spectral threshold values. The developed rule set was transferred to secondary scenes with minimal changes to the master rule set. Visual interpretation was used to validate results. The results showed high agreement of absolute numbers for automatically extracted dwellings (15 349) versus visually extracted dwellings (14 261).

Shekhar (2012) delineated informal settlements in Pune City, India, using QuickBird imagery. The study highlighted the efficacy of the developed methodology to discriminate informal dwellings by describing typical characteristics of these settlements. Fuzzy membership functions

of texture, geometry, and contextual information were used to achieve an overall accuracy of more than 87%.

Kohli et al. (2011) expanded upon the work of Hofmann et al. (2008) and developed Generic Slum Ontology (GSO), which can be used as part of a conceptual classification OBIA schema. Kohli et al. (2012) followed an ontological approach to conceptualise informal settlements using class indicators of the built environment in Ahmedabad, India.

Kohli et al. (2013) used textural features such as entropy and contrast derived from a grey level co-occurrence matrix (GLCM) combined with an adapted GSO to identify informal settlements in Ahmedabad, India using GeoEye-1 imagery. The OBIA-based classification was applied to three different subsets with minimal adaption and achieved final accuracies ranging from 47% to 68%. The results showed that visually different urban patterns could be classified using a combination of different texture features to increase classification accuracy

2.9 Review of building extraction techniques

The availability of VHR imagery has driven the demand for high resolution digital elevation models (DEM) that can be used to provide the geospatial information required for the identification and planning of informal settlements (Krauβ & d'Angelo 2011; Kit et al. 2012; Kohli et al. 2013). A DEM is a digital representation of the earth that can be used to represent the surface or terrain, and is referred to as a digital surface model (DSM) or digital terrain model (DTM). The two most common and well-established sources for digital elevation modelling include light detection and ranging (LiDAR) and stereo photogrammetry (Demir, Poli & Baltsavias 2009).

2.9.1 LiDAR

LiDAR is an active sensor that emits laser pulses in the form of high energy particles (photons) towards the earth's surface with a pulse repetition frequency (PRF) (Mallet & Bretar 2009). The time taken for the photons to reflect back (referred to as returns) to the photodetector is recorded. LiDAR provides high-resolution vertical and horizontal spatial accuracy and has become an important and primary data source for generating DEMs (Moreira et al. 2013), more specifically in the built urban environment (Remondino et al. 2014).

Several different varieties of LiDAR exist, i.e. terrestrial, bathymetric, atmospheric, space borne and airborne. Airborne LiDAR used in this study consists of three separate technologies that help to ensure accurate data collection. The laser transmits and collects the laser pulses, the global positioning system (GPS) provides position information so that collected points can be

referenced to the earth's surface, and the internal measurement unit (IMU) tracks the altitude of the aircraft recording changes in roll (x-axis), pitch (y-axis), and yaw (z-axis). Huising and Pereira (1998) outlined three main sources of error as a result of 1) laser pulse delay, 2) GPS misalignment, and 3) errors arising from system calibration.

Flood (2002) identified five levels of LiDAR data provided by vendors that are based on the needs of the users applications namely; 1) basic or 'all-points', 2) low fidelity or 'first-pass', 3) high fidelity or 'cleaned', 4) feature layers, and 5) fused. LiDAR data from levels 1 through 3 are processed at increasing degrees of data filtering but have no feature identification, whereas levels 4 and 5 consists of extracted features and consequently incur higher costs and longer delivery time (Cheuk & Yaun 2009).

LiDAR pulses are recorded in two ways either discrete return (DR) systems, or full-waveform (FW) system. DR systems work by sending out pulses and collecting the number of individual returns. This means that the laser might initially hit an object and part of it will return while some will pass through the object and return when it hits the next object. Earlier LiDAR systems recorded only one return, either first return or the final return in the reflective wave (Jensen 2007). Since 2000, commercial systems are capable of measuring multiple returns per pulse. In an urban environment, by assessing the return signals, the difference between buildings, vegetation, and terrain can be distinguished. The first return is representative of features above the ground surface, such as buildings, bridges, and tree canopies and can be used to create digital surface models (DSM). Intermediate returns are helpful in separating vegetation from objects above ground, whereas final returns are the first approximation of the bare ground and are used to create digital terrain models (DTM). FW systems work differently, as incoming pulses are not recorded as returns, but rather the amount of energy returned to the sensor which is measured over a period of time at equal intervals, and is referred to as intensity.

Ussyshkin and Theriault (2011) present an extensive review on DR and FW LiDAR systems, and note that each collection mode has distinct advantages and disadvantages that vary depending on the application. Furthermore the authors acknowledge the advances in DR LiDAR technology and its ability to capture high-quality dense point clouds for automated modelling and analysis. The availability of FW and DR systems provides the opportunity to represent urban structures with high vertical and horizontal spatial accuracy, and has resulted in LiDAR becoming an important and primary data source for generating detailed DEMs.

Bujan et al. (2013) evaluated the effect of point density on classification accuracy using multitemporal and multidensity LiDAR data. It was found that certain classes were unable to be

discriminated when using densities below 4 pulses/m^2. The use of low density LiDAR decreased the quality of information and led to variations in the average values of the data. Nonetheless, the reduction of LiDAR point density only slightly affected the results as densities below 4 pulses/m^2 achieved an overall accuracy of 85% and above achieved 93%.

Hermosilla (2011) presented a quality assessment of two approaches for automatic building detection and localisation using high spatial resolution imagery and LiDAR data. The methodological parameters were adapted to three study areas: urban, suburban and industrial buildings. The results (97%, 95% and 98%) show high efficiency for building detection using an nDSM when parameters are adjusted to the type of urban landscape under consideration. LiDAR was capable of capturing high vertical accuracy; however, it was found that LiDAR-derived products may contain uncertainty in complex urban environments. Uncertainty was caused by reflective surfaces and their corresponding albedos (Böhler, Bordas, and Marbs 2003), as well as when tall trees grow adjacent to buildings. To decrease uncertainty and increase planimetric accuracy, Awrangjeb and Fraser (2013) recommend the integration of imagery with LiDAR-derived products.

2.9.2 Stereo photogrammetry

Stereoscopic photogrammetry is based on the phenomenon of how the brain interprets depth perception. This is possible due to the physical distance between the eyes that provides the brain with slightly different viewing angles of the same scene (Mallet et al. 2011). An equivalent artificial setup can be achieved by taking imagery of the same object or scene from different viewing angles, allowing the observer to perceive height and depth in overlapping photographs (Mallet et al. 2011). Traditionally, a scanning stereoscope was used to look at overlapping photography or stereo-pairs simultaneously (Church 1948); however, with the advent of digital imaging more research was focused on automated procedures to replace manual operator systems (Gruen 1985). The 1990s was a time of consolidation for image matching (Remondino & Menna 2014) and a large number of commercially available photogrammetric platforms became available. The process of deriving DEMs can be summarised into four main steps namely 1) image acquisition, 2) aerial triangulation, 3) image matching, and 4) interpolation.

2.9.2.1 Image acquisition

Traditional stereo photogrammetry principles can introduce distortions during image capture due to radial distortions, refraction, the curvature of the earth, focal length, and perspective effects (Shenk 2005). In order to identify and correct distortions, camera calibration is typically provided by the vendor or manufacturer, which is used to compute and correct interior

orientation. (LPS Project Manager User Guide 2009). A camera with known interior orientation that does not change is referred to as a metric camera (Shenk 2005).

During image capture, the imagery can be prone to geometric and atmospheric errors that need to be corrected. Atmospheric attenuation/scattering can be corrected by converting radiance to reflectance or by the application of radiometric normalisation. Sources of geometric distortion include 1) instrument error, 2) platform instability, and 3) and relief displacement. These errors can be corrected using rectification methods such as registration (register one image to another), geo-referencing (assign map coordinates to image), or orthorectification (correct for terrain relief displacement) with the aid of ground control points (GCPs), tie points, and existing elevation information. To ensure high-quality surface reconstructions, a horizontal image overlap of 60% and vertical image overlap of 20% per area are recommended to allow evenly distributed GCPs and tie points. An inadequate horizontal and vertical image overlap can result in low-quality surface reconstruction and the failure of image matching (Geomatica Orthoengine User Guide 2009).

2.9.2.2 Aerial triangulation

Aerial triangulation represents the mathematical process of establishing a precise and accurate relationship between the image (camera position) and the ground using GCPs and tie points to ensure accurate image and model orientation. Not all GCPs and tie points collected will have the same reliability as they will be weighted inversely to their estimate error (Geomatica 2013). This coordinate information is then used to calculate exterior information (x, y, z coordinates and three rotation angles Omega, Phi, and Kappa).

There are four main steps for aerial triangulation namely 1) preparation, 2) point transfer, 3) point measurement, and more importantly 4) block adjustment. Aerial triangulation is often referred to as block adjustment, and is the process of adjustment (transformation) that results in orthorectified image mosaics.

2.9.2.3 Image matching

After the interior information (camera calibration) and exterior information (aerial triangulation) have been determined, image matching (IM) can be applied to the overlapping (stereo) images. IM refers to the process of finding matching points (conjugate points) in corresponding stereo image pairs, computing 3D positions in object space, and assessing the quality of the matching (Shenk 1999).

There are two common types of IM algorithms used in stereo photogrammetry (Joglekar & Gedam 2012), namely:

- Area-based matching
  - Local methods
  - Fourier methods
  - Cross-correlation
- Feature-based matching
  - Dreschler
  - Moravec
  - Förstner

In area-based matching, the grey level of a small area (search window) of an image is compared with the corresponding matching entity in the other image (Vossleman et al. 2004). Fourier methods are robust methods for highly correlated and frequency dependent noise disturbances in varying conditions in the image scene (Joglekar & Gedam 2012). Local methods have been found to give results with high speeds and produce better results than Fourier methods (Joglekar & Gedam 2012). Cross-correlation (CC) methods are sensitive to changes in intensity and sensor type and are best suited for real-time applications. There are, however, limitations to CC methods in terms of the high computational complexity and similarity measures within the image scene (Joglekar & Gedam 2012). In feature-based matching, features entities such as points, edges and patches are independently extracted in all the images, and then matched with corresponding feature entities of other images by comparing feature attributes (Joglekar & Gedam 2012). Feature-based matching can be strongly influenced by three features within an image namely 1) discreteness, 2) similarity, and 3) consistency (Harika, Rao & Krishnaiah 2013). Features such as points can distinguish distinct objects in the image and are referred to as discreteness, whereas the resemblance of two points is known as similarity. Lastly, feature-based matching can be influenced by the consistency of the surface in the images and is referred to as the conformity of the features (Harika, Rao & Krishnaiah 2013). In addition, Lowe (2004) notes that low contrast or homogeneity within an image can constrain the feature-based method to track features across the image. Area-based methods have been found to outperform feature-based methods as they are computationally more efficient, comparatively more active, and take into account a whole neighbourhood around the points to establish correspondence (Joglekar & Gedam 2012).

2.9.2.4 Interpolation

Surface interpolation is used to predict values of cells at locations that lack sampled points based on the principle of spatial autocorrelation, which measures the degree of dependence between near and far things (Child 2004). Interpolation methods can produce varying predictions as each model makes different assumptions of the data and uses different calculations (Child 2004).

There are two categories of interpolation techniques: deterministic and geospatial/stochastic. Deterministic interpolation techniques create surfaces based on measured points or mathematical formulas and include methods such as inverse distance weight (IDW), natural neighbour (NN), and Spline. IDW determines cell values using a linear-weighted combination set of sample points such that the greater the distance, the less influence (weight) the cell has on the output cell value (Garnero & Godone 2013). Splines is a technique that minimises overall surface curvature by creating a smooth surface that passes through all input point, and as a result is best suited for smoothly varying surfaces (Child 2004). NN interpolation finds the closest subset of input points, and applies a weight on proportionate areas to interpolation unknown values (Sibson 1981). Garnero and Godone (2013) noted that no artefacts such as peaks, pits, or ridges will be generated in the data as NN infers no trends about the data and is well suited for clustered scattered points. Geospatial techniques are based on statistics and are used for more advanced prediction modelling and provide probabilistic estimates. Kriging is a powerful statistical method that assumes that the distance between sample points reflects spatial autocorrelation and can be used to explain surface variations (Child 2004). A more extensive review of spatial interpolation techniques and corresponding formulas can be found in the *Principles of Geographical Information Systems* by Burrough, McDonnel and Lloyd (2015)

Stereoscopic photogrammetry has been successfully applied in several studies for 3D building extraction using stereo satellite (Aguigiaro, Poli & Remondino 2012; Yan et al. 2013; Qin 2014) or stereo aerial imagery (Salehi et al. 2011; Tarantino & Figoritto 2011; El Garouani et al. 2014). For example, Nex and Remondino (2012) extracted roof outlines (eaves, ridges, and pitches) from stereo aerial images and compared derived point clouds to LiDAR. The authors note that IM techniques can provide dense and reliable point clouds, comparable to LiDAR in accuracy and the level of detail that can be achieved. Moreira et al. (2013) evaluated image matching using stereo GeoEye-1 and aerial imagery against LiDAR to generate a DSM for automated building extraction. DSMs were compared on both planimetric and height information using reference building dimensions. The authors found that the percentage of correctly reconstructed buildings from the different DSMs in the test areas were very similar. The highest percentage of correctly reconstructed models was LiDAR (67%), followed by aerial (59%), and GeoEye-1

(41%). LiDAR achieved the lowest mean residual of 0.64 m, followed by aerial (1.82 m), and GeoEye-1 (1.94 m). Satellite imagery did not provide DSMs that were suitable for the automatic extraction of detailed buildings, whereas both aerial imagery and LiDAR provided accurate and detailed alternatives. More recently, Poli et al. (2015) investigated the radiometric and geometric characteristics of three VHR optical sensors, namely GeoEye-1, WorldView-2, and Pleaiades-1A for urban feature extraction. DSMs were generated from stereo pairs and were also quantitatively assessed using a LiDAR DSM as a reference. The results for all three image datasets indicate that the root-mean-square-error (RMSE) ranged from 6.1 m to 8.5 m when compared the LiDAR DSM.

2.9.3 Structure from Motion

The introduction of commercially available unmanned aerial vehicles (UAVs) and significant advancements in hardware and algorithms have seen photogrammetry re-emerge as a competitive technology in the form of structure from motion (SfM). The principles of SfM developed from an overlooked computer vision technique (Ullman 1979; Spetsakis & Aloimonos 1991; Boufama et al. 1993) to an emerging methodology used to estimate the 3D geometry of features from a collection of 2D imagery in model space (Snavely et al. 2007; Snavely 2008; Agarwal et al. 2010; Furukawa & Ponce 2010). Like stereo photogrammetry, SfM can be summarised into four steps, namely 1) feature matching, 2) key-points to 3D point cloud, 3) point cloud densification, and 4) 3D meshing and DSM creation.

2.9.3.1 Feature matching

Features and key points within the images are identified using the scale invariant feature transformation (SIFT) algorithm (Lowe 2004). The four major stages of the SIFT are: 1) scale-space extrema detection, 2) keypoint localisation, 3) orientation assignment, and 4) key point descriptors (Lowe 2004). SIFT determines key points of interest in the image by searching image scales and locations using the difference-of-Gaussian function (Lowe 2004). This allows candidate image locations to be created and an orientation to be assigned to each location based on the local gradient direction. This provides invariance to the transform as the scale, orientation, and location are processed for feature matching (Mikolajczyk & Schmid 2002). A number of alternatives to SIFT exist, such as speeded up robust features (SURF), LDAHash, the principle component analysis (PCA) SIFT, gradient location, and the orientation histogram. These methods all aim to achieve the same results, however, SIFT has proven to be the most robust to large image variations (Harwin & Lucieer 2012). A number of software solutions use feature-matching methods, including, Microsoft Photosynth (Mircosoft 2010), Agisoft PhotoScan (Agisoft 2010) and PhotoModeler (Eos Systems 2011). These commercial applications, however,

can take on ‘black box’ characteristics with proprietary blend and alterations to existing algorithms, resulting in the user needing to have no prior or expert knowledge but offer less control over the quality of the output and the management of errors (Machletti et al. 2015).

2.9.3.2 Key points to 3D point clouds

Matching key-point descriptors between image pairs is achieved by using the method of Arya et al. (1998) for approximating nearest neighbour kd-tree while maintaining Lowe’s ratio test (Lowe 2004). A 3D model and sparse point cloud of the object are produced by using an iterative process called random sampling consensus (RANSAC) to calculate epipolar geometry (Bolles & Fischler 1981). RANSAC finds corresponding key points across multiple image pairs producing a connectivity graph while representing incorrectly estimated key-points as outliers (Furukawa & Ponce 2007; Furukawa et al. 2010). RANSAC is further used to calculate and optimise camera positions using a Direct Linear Transform (DLT) (Bolles & Fischler 1987), and no prior knowledge regarding exterior orientation is needed (Snavely et al. 2008).

The success of matching descriptors is based on the number of key-points in an image, and is primarily dependent on image texture, the density of points, sharpness, the number of images used (minimum three), and the resolution of the imagery (Westoby 2012). More recently, Westoby (2012) found that increasing the number of overlapping images used to find key-points, resulted in significantly increased processing time.

2.9.3.3 Point cloud densification

There are several approaches to computing a 3D mesh from a large set of objects derived from point clouds including plane sweeping-based, stereo-based, or growing-based methods (Jancosek & Pajdla 2011). Similarly, the Centre for Machine Precision’s multi-view stereo (CMPMVS) is a multi-view reconstruction approach that uses the Delaunays tetrahedralization adapted from the Labatut CGF 2009 method to identify point occlusion within images and derive points clouds of weakly supported surfaces (Jancosek & Pajdla 2011). CMPMVS has gained in popularity and has shown superiority to other methods in its ability to reconstruct difficult surfaces while preserving details of the same time quality and computational time (Jancosek & Pajdla 2011).

Anomalies can introduce problems when performing disparity estimation and point cloud rendition, such as the presence of lens flares, high specular reflection from cars and windows, homogeneous texture of the road surface, and repetitive patterns on facades in an urban scene (Cornelis et al. 2008). A number of methods have been proposed to address the difficulties in the disparity estimation process, including 1) geometric constraint, 2) stereo camera rectification, 3)

similarity measures, 4) line selection, 5) inter-frame smoothing, and 6) general stereo camera rectification. A comprehensive review of the above proposed methods for anomaly/object recognition in images can be found in Cornelis et al. (2008). More recently, NVidia has developed a high-definition 3D Graphic Processor Unit (GPU) called the compute unified device architecture (CUDA) that has enhanced computational power and memory bandwidth for SfM based applications (CUDA 2013). This has resulted in software providers (both proprietary and some open source) using this technology, as it has been found that GPU-based processing significantly decrease computational time (Hirschmuller et al. 2012).

2.9.3.4 3D meshing and DSM creation

The 3D mesh or DSM generated is generally in an arbitrary reference frame and needs to be to be registered to a real-world coordinates system (Harwin & Lucieer 2012). This is achieved by identifying key features within the 3D mesh/DSM, which can be matched to real-world features using GCPs. Once features have been established and matched, 3D Helmert transformations can be used to transform the data to real-world coordinates along seven parameters: three translations, three rotations, and one scale (Harwin & Lucieer 2012). It is important to note that geo-referencing can be undertaken at varying stages of image processing.

SfM has been reviewed in various urban applications such as the 'Building Rome in a day' project (Agarwal et al. 2009), progress of building erections (Golparvar-Fard et al. 2011), and the detection of urban facades (Ceylan et al. 2014). Mathews and Jensen (2014) demonstrated urban building extraction and reconstruction using SfM. It was found that the generated point cloud accurately represented the features of interest with high point densities throughout the model. The authors noted that the increased model detail enabled accurate height and distance measurements to be performed. The Literature suggests that the success of SfM reconstruction is a function of spatial resolution and density of image points that are defined by altitude and field of view (FOV) of image capture (Javernick et al. 2014). A study by Remondino et al. (2014) revealed the differences in point cloud densities that can be achieved by image matching, SfM, and LiDAR. A typical point cloud density of a LiDAR dataset can be in the range of 1 to 25 points/m^2, whereas aerial photogrammetric imagery with a ground sampling distance (GSD) of ~10 cm using image matching can theoretically produce point cloud densities of up to 100 points/m^2, however, SfM with the same GSD can produce point cloud density greater than 125 points/m^2, depending on the altitude and FOV used.

CHAPTER 3: EXTRACTION AND EVALUATION OF NDSMS FROM LIDAR, PHOTOGRAMMETRIC IMAGE MATCHING AND STRUCTURE FROM MOTION FOR INFORMAL SETTLEMENT MAPPING IN CAPE TOWN*

3.1 Abstract

Three approaches to extract nDSM models were investigated: (i) light detection and ranging (LiDAR) data, (ii) high resolution aerial photographs in a process of image matching, and (iii) a series of aerial images captured using a hand-held camera using structure from motion (SfM) techniques. SfM is a novel technique that has not been widely used for nDSM extraction. This study represented a first attempt at evaluating the three approaches, particularly for mapping informal dwellings. The accuracy of the respective nDSMs was evaluated using vertical profiles, area-based, and positional-based accuracy assessment metrics. In the train site, the highest accuracy was achieved LiDAR, followed by SfM and image matching in the vertical profiles (2.39 m, 2.54 m, and 3.14 m) and positional-based accuracy (96.33%, 91%, and 86.65%). Completeness and correctness was also recorded, with LiDAR achieving the highest completeness (85%), followed by SfM (82%), and image matching (68%). Whereas, image matching achieved the highest correctness (93%), followed by SfM (79.5%), and LiDAR (74.6%) To further test the robustness of the nDSMs, each model was applied to an independent test site with varying dwelling arrangements. LiDAR achieved the highest vertical profile accuracy and positional-based accuracy (2.08 m and 97.17%), followed by image matching (2.10 m and 87.38%). and SfM (2.14 m and 93.46%). Overall LiDAR achieved the highest accuracy in all three accuracy assessments, showing its ability to handle the complex and undefined morphology of the informal settlements.

3.2 Introduction

Obtaining up-to-date geospatial information for mapping, monitoring, and visualization of informal settlements is one of the major challenges to modern remote sensing in the urban environment (Kit et al. 2013; Kohli et al. 2013). The dynamic spatial-temporal behaviour, the

---

*This chapter will be submitted for publication and will conform to the prescribed structure of that journal. Some of texts used in this chapter are identical to those of previous chapters, as the same data and methods were used.

relatively high inner-structural heterogeneity (Hoffman et al. 2008; Pinho et al. 2011; Shekhar 2012), and microstructure of informal settlements do not allow generically applicable urban solutions to be applied (Kohli et al. 2013; Taubenbock & Kraff 2013). Remote sensing (RS) and advanced image processing methods enable urban mapping and automated extraction of buildings which can provide predictable and consistent results for data collection, methodology development and analysis (Hoffman et al. 2008; Pinho et al. 2011; Shekhar 2012). The availability of very high resolution (VHR) imagery have driven the increased demand for high resolution digital elevation models (DEM), that can potentially be used to fulfil the geospatial information required for the identification and planning of informal settlements (Krauβ & d'Angelo 2011; Kit et al. 2013; Kohli et al. 2013).

The two most common and best established sources for topographic mapping and digital elevation models are light detection and ranging (LiDAR), and photogrammetric image matching (IM). During the last 30 years LiDAR has become a reliable and fundamental source for dense point cloud generation (Remondino et al. 2014). LiDAR is an active sensor that emits laser pulses in the form of high energy particles (photons) towards the earth surface with a pulse repetition frequency (PRF) (Mallet & Bretar 2009). The time taken for the photons to reflect back (referred to as returns) to the photodetector is recorded. LiDAR provides high-resolution vertical and horizontal spatial accuracy and has become an important and primary data source for generating DEMs (Brennan & Webster, 2006; Moreira 2013). IM is powerful tool for topographic modelling and are based on the principles of stereoscopic photogrammetry. Stereoscopic photogrammetry uses series of overlapping images to find corresponding pixels in stereo pairs by reconstruction called triangulation. Error is reduced in photogrammetry by defining sensor calibration and image orientation which describe the inherent camera geometry, lens distortion and parallax (Westoby et al. 2012). The introduction of commercially available unmanned aerial vehicles (UAVs) and the significant advancements in hardware and algorithms, have seen photogrammetry 're-emerge' (Westoby et al. 2012:2) as a competitive technology in the form of structure form motion (SfM) (Westoby et al. 2012). The principles of SfM developed from a computer vision technique (Spetsakis and Aloimonos 1991; Boufama et al. 1993) to an emerging methodology used to create three-dimensional point cloud data in model space (Agarwal et al. 2010; Furukawa and Ponce 2010; Snavely 2008; Snavely et al. 2006). SfM operates under the same tenets as stereoscopic photogrammetry, but unlike photogrammetry that uses overlapping stereo images, SfM uses three or more images in a Multiview-stereo approach to accurately calculate distances between the camera and objects (McGlone & Lee 2013; Furukawa and Ponce 2010; Vu et al. 2012; Stal et al. 2013). In addition, SfM supports changes in

camera perspective and image scale using a highly redundant bundle adjustment method based on matching features (Westoby et al. 2012). This eliminates the need for orientation parameters or the location of ground control points (GCP) to be defined prior to processing (Lowe 2004; Snavely et al. 2008). SfM has been notably popularised through a range of cloud-processing engines including Microsoft Photosynth, Bundler, Agisoft Photoscan, and PhotoModeler.

Several studies have investigated the ability of SfM to generate comparable models to LiDAR and IM in terms of point cloud density. Westoby et al. (2012) and Fonstad et al. (2013) found that SfM point clouds was comparable to both IM and LiDAR point clouds, however SfM has the potential to produce denser point clouds than LiDAR (Nex & Remondino 2012) and IM (Leberl 2010;Remondino et al.2014). A recent study by Remondino et al. (2014) revealed the differences in point cloud densities that can be achieved by IM, SfM and LiDAR. The authors reported that a typical point cloud density of a LiDAR dataset can be in the range of 1 to 25 points/m^2, whereas aerial photogrammetric imagery with a ground sampling distance (GSD) of ~10 cm using image matching can theoretically produce point cloud densities of up to 100 points/m^2, however SfM with the same GSD can produce point cloud density greater than 125 points/m^2. In addition Javernick et al. (2014) found that the density of the point clouds generated is however dependent on the spatial resolution (i.e. GSD), which is defined by the altitude and field of view (FOV) of image capture, such that an increased GSD can result in an increased point cloud density. Several authors have investigated the utility of IM, SfM and LiDAR for purposes of planning and modelling in urban areas. LiDAR has become an influential dataset to characterise urban features (MacFaden et al. 2012; O'Neil-Dunne et al. 2012; Aguilar, Saldaña, & Aguilar 2012; Hamedianfa et al. 2014), building modelling and reconstruction (Awrenjeb and Fraser 2012; Sun & Salvaggio 2013), and aid building footprint extraction (Rutzinger et al. 2010; Awrenjeb & Fraser 2012; Brédif et al. 2013). For example, Hermosilla (2011) showed the efficiency of LiDAR derived nDSMs for building detection. The author adapted the methodological parameters to the respective study areas and achieved good results in urban (97%), suburban (95%) and industrial (98%) areas. IM has been successfully applied in several studies for 3D building extraction using stereo satellite (Aguigiaro, Poli & Remondino 2012; Yan et al. 2013; Qin 2014) or stereo aerial imagery (Salehi et al. 2011; Tarantino and Figoritto 2011; El Garouani et al. 2014). Poli et al. (2015) investigated the radiometric and geometric characteristics of three VHR optical sensors namely GeoEye-1, WorldView-2, and Pleaiades-1A for urban feature extraction. DSMs were generated from stereo pairs and were assessed quantitatively using a LiDAR DSM as a reference. The results for all three image datasets indicate that the root-mean-square-error (RMSE) ranged from 6.1m to 8.5m when compared the

LiDAR DSM. SfM have been reviewed for various urban applications such as the 'Building Rome in a day' project (Agarwal et al. 2009), progress of building erections (Golparvar-Fard et al. 2011) and the detection of urban facades (Ceylan et al. 2014). More recently, Mathews & Jensen (2014) demonstrated urban building extraction and reconstruction using SfM. It was found that the generated point cloud accurately represented the features of interest with high point densities throughout the model. The authors noted that the increased model detail enables accurate height and distance measurements to be performed.

A limited number of studies have used IM, LiDAR and SfM in combination. Moreira et al. (2013) evaluated IM using stereo GeoEye-1 and aerial imagery against LiDAR to generate a DSM for automated building extraction. DSMs were compared on both planimetric and height information at 1:1000 scale using reference building dimensions. The authors found that the percentage of correctly reconstructed buildings from the different DSMs in the test areas were very similar. The highest percentage of correctly reconstructed models was LiDAR (67%), followed by aerial (59%), and GeoEye-1 (41%). LiDAR achieved the lowest mean residual of 0.64m, followed by aerial (1.82m), and GeoEye-1 (1.94m). It was found that satellite imagery did not provide DSMs that were suitable for the automatic extraction of detailed buildings, while both aerial imagery and LiDAR provided accurate and detailed solutions. In a related study, Stal et al. (2013) compared DSMs extracted from stereo aerial imagery using IM and LiDAR data. It was found that, although LiDAR and IM are different measurement systems, the results are highly qualitatively and quantitatively comparable in terms of the degree of detail and coverage size. The models were compared relative to each other by pixel-wise differencing of DSM values and a set of random reference points. Based on the calculated statistics, there were no significant elevation differences. The authors found that large blunders and noise can be observed during the IM process and that the projection can invariably cause distortion and image loss. Furthermore the lack of quality and insufficient detail of the model was observed to have a significant impact on accuracy and performance. A multiscopic or multiview-stereo approach was recommended to reduce the effects of occlusions and distortions in DSM models.

Andrews, Bedford and Bryan (2013) compared LiDAR and SfM derived DSMs of an English historical site. The accuracy of the DSMs was evaluated using variance, standard deviation, RMSE, and coverage. The authors found no significant accuracy variations between the two datasets. More recently Maiellaro, Zonno and Lavalle (2015) compared a SfM point cloud using terrestrial imagery with a laser scanner point cloud mounted on an UAV. The point clouds were compared using the open source software CloudCompare that computed maximum deviation. It was observed that point deviation between the datasets ranged from less than a centimetre and up

to 6cm in certain places, and that terrestrial imagery compared favourably to laser scanner datasets. Likewise Kersten, Mechelke and Maziull (2015) compared outputs of different SfM models and terrestrial laser scanners (TLS) but contrary to the finding of Maiellaro, Zonno and Lavalle (2015), the authors noted that the geometric quality of the derived point clouds and surface models were significantly different, and that SfM could not achieve comparable accuracy to the laser scanner. This paper evaluates SfM, IM and LiDAR for urban building extraction in informal settlements. To achieve this, 1) an nDSM was extracted using established methods of IM and LiDAR and compared to an nDSM generated using SfM and high resolution datasets, and 2) the resulting nDSMs were evaluated using vertical profiles, area-based and positional-based accuracy assessment metrics. To assess the robustness of the models, it was applied to a secondary independent test site with different informal settlement spatial structures. A workflow of the methods used in this chapter is presented in Figure 3.1.

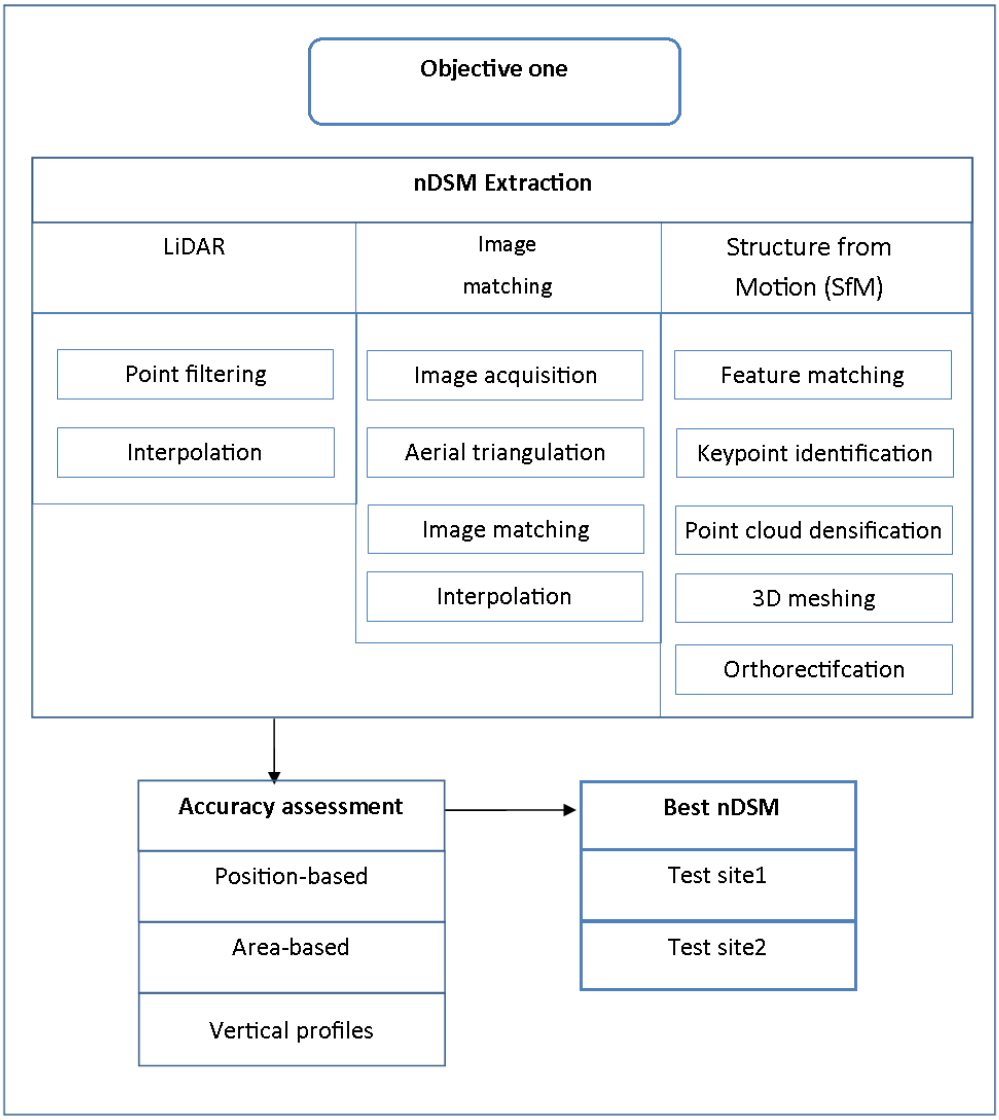


Figure 3.1 Workflow of the methods used in chapter 3. The best nDSM will be used as ancillary classification data in chapter 4.

3.3 Methods and materials

3.3.1 Study sites

The study area seen in Figure 3.2 is situated in the Western Cape, South Africa. Two different types of unstructured informal settlements were chosen for this study. Test site 1 has high density, very crowded and clustered dwellings, whereas test site 2 is a relatively open area with sparse dwellings and low density. Majority of the dwellings are single storied structures, display simple geometry (4-sided), constructed from diverse materials with variable texture and colours, and with separation distance of ~1 m between dwellings (Mason and Baltsavias 1997).

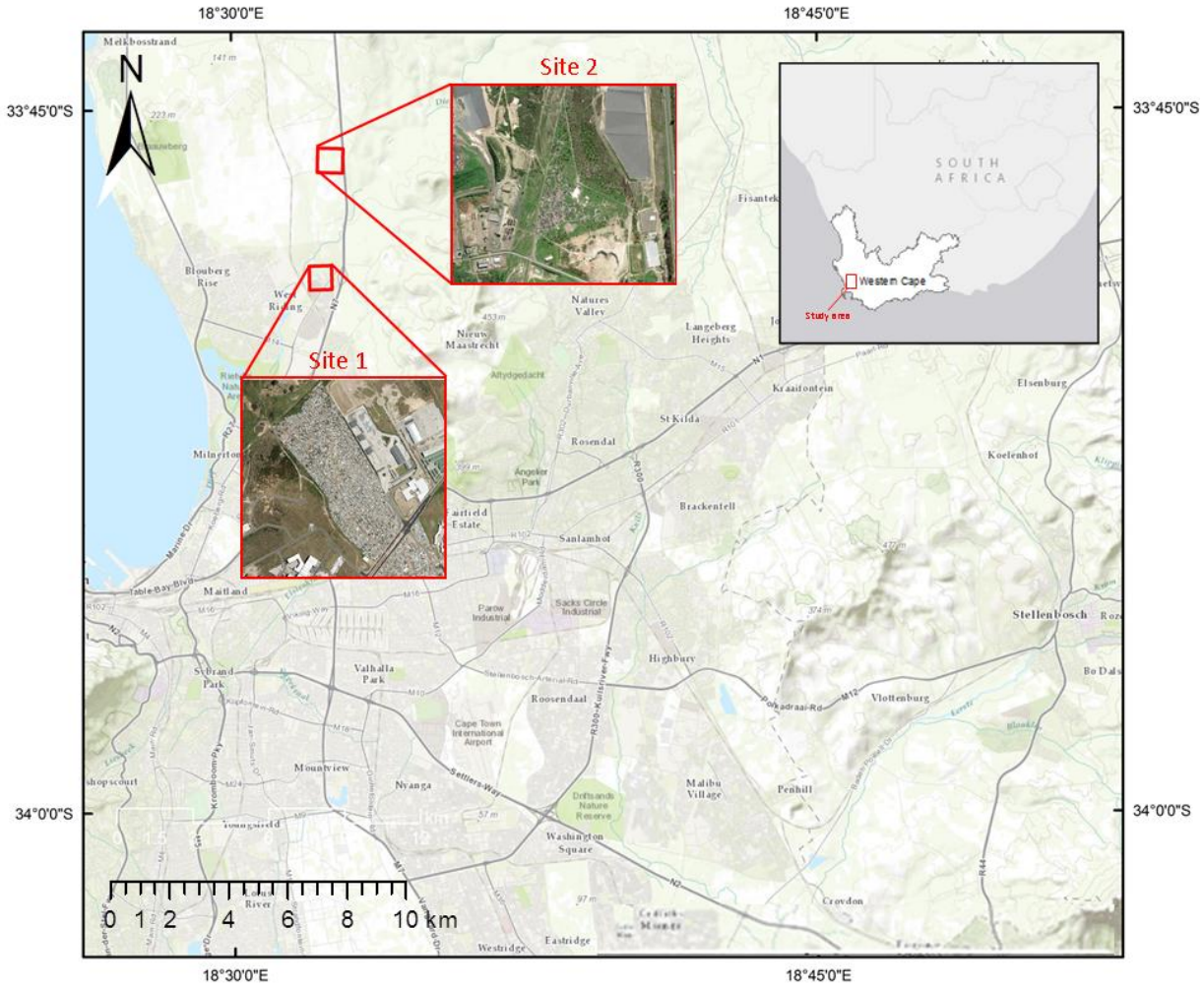


Figure 3.2 Location of the study area with WorldView-2 inset overlaid on ESRI ArcMap imagery base map.

3.3.2 Image and field data

Very high resolution (VHR) (12.5 cm) orthorectified digital aerial imagery and raw stereo-imagery for test site 1 and test site 2 were captured in December 2013, and was acquired from City Maps, City of Cape Town. These photographs were captured using Microsoft UltraCam Eagle that recorded the exterior (X, Y, Z) and interior (ω, ϕ, κ) orientation of each digital photograph.

LiDAR was acquired from the Geospatial Services, City of Cape Town and was captured in December 2013 using the ALS50 Airborne Laser Scanner developed by Leica Geosystems based on the Fuguro Fli_Map system. It was recorded in a sinusoidal scan pattern, nominally centred about nadir, achieving an average point density of 1 point/m^2.

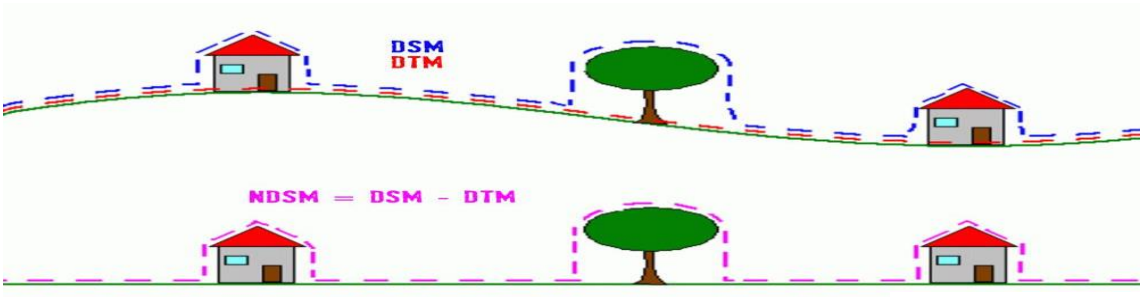
The SfM image dataset was captured using a hand held Canon D600 taken from a gyrocopter (aircraft). The flying height ranged from 270-510 m achieving a ground sample distance (GSD) of 6.5-12 cm. The specific flying height was selected to maintain a GSD comparable to a GoPro Hero3 used on commercial low altitude UAVs flown at approximately 80-120m. A gyrocopter was used as a substitute due to current UAV legislative restrictions in South Africa, resulting in this comparable alternative.

The SUDEM was developed by the Centre for Geographical Analysis (CGA) at Stellenbosch University, South Africa. A combination of interpolation algorithms was used on large scale (1:10,000) contours, spot heights and the SRTM DEM to interpolate a 5m DEM product (Van Niekerk 2012). The SUDEM (2015) was used as an auxiliary DTM dataset for nDSM creation. The very high resolution (12.5 cm) orthorectified aerial imagery was printed on A3 hard copy maps and was used to identify and verify informal dwellings in the field.

3.3.3 Data processing

3.3.3.1 Normalised digital surface model generation

The focus of this study is to generate nDSM which represents absolute elevation of objects above the DTM. In the context of this study, absolute heights of dwellings above ground were needed. The subtraction of the DTM from the DSM results in absolute height values of objects as shown in Figure 3.3.



Source: DLR (2013)

Figure 3.3 The graphic representation of an nDSM.

3.3.3.2 LiDAR

The raw LiDAR point data was pre-processed by Geospatial Services into ground and non-ground points, and was provided in a binary format (LAS) that included echoes of the pulses, GPS time and intensity. The LAS files were then processed using the 'LAS dataset to raster tool' in ESRI ArcMap 10.2.1. The natural neighbour method of interpolation was used to find the closest subset of samples to a point, and applies weights based on proportionate areas to interpolate a value (Sibson 1981). The DSM was derived by interpolating the pre-processed ground and non-ground points together, whereas the DTM was derived by only interpolating ground points. Subsequently an nDSM was derived using the formulae shown in Figure 3.3.

3.3.3.3 Image matching

The stereo pair of aerial images was processed using the photogrammetric software package OrthoEngine in Geomatica 2013, developed by PCI Geomatics. The (Rigorous) Aerial photography math model was used to calculate the position and orientation of sensor at the time of image capture (PCI Geomatics 2013). The correct camera calibration information such as focal length, principal point offset radial lens distortion and chip sizes was used to identify and correct image distortions. The rational polynomial coefficient files (RPC) was provided by the aerial imagery vendor, and contains third degree polynomial coefficients that relate the sensor's internal geometry and external geometry at the time of image capture. GCPs and tie points were captured on stereo pairs to supplement exterior orientation provided by the RPC. Bundle adjustment was computed to accurately calculate the position and orientation of the sensor. The resulting DSM was extracted and the nDSM was subsequently computed by subtracting the SUDEM (DTM) from the DSM. In Figure 3.4, the underlying principle of bundle adjustment is illustrated. After matching pixels in the two (stereo) images are extracted, the sensor geometry from the computed math model is used to calculate x, y, and z positions.

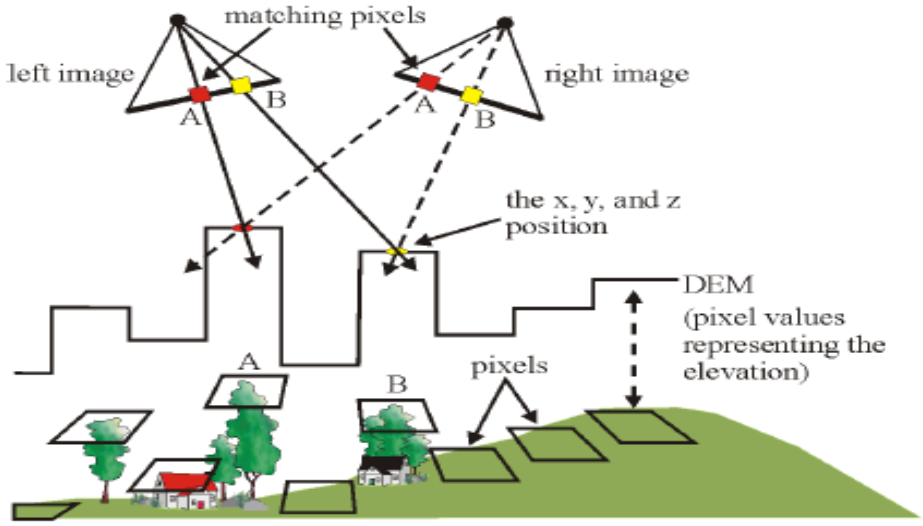
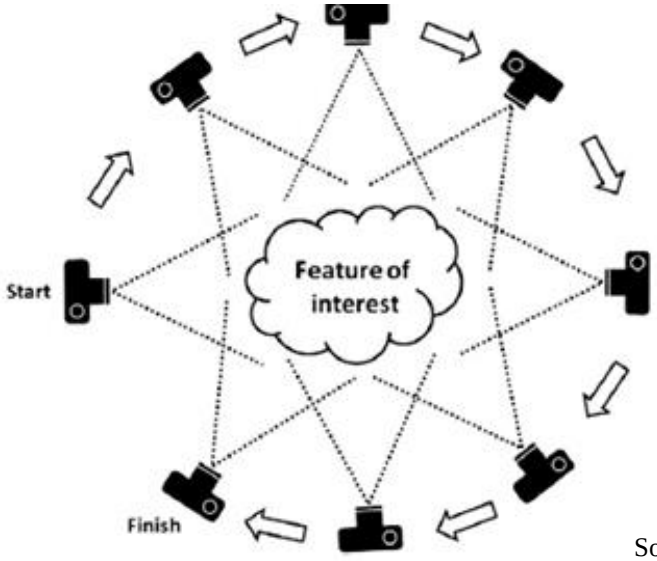


Figure 3.4 Underlying principle of bundle adjustment in OrthoEngine. Source: PCI Geomatica (2013)

3.3.3.4 SfM

A DSM was created using the gyrocopter imagery as input in the Designing Reality SfM software. Unlike stereo photogrammetry that uses overlapping stereo images, SfM uses three or more images in a MVS approach to accurately calculate distances between the camera and objects (Furukawa & Ponce 2010; Vu et al. 2012; Stal et al. 2013; McGlone & Lee 2013). The basic workflow produces two datasets, a sparse point cloud and a dense point cloud. The generation of the SfM DSM can be described in three simplified steps.



Source: Westoby (2012)

Figure 3.5 Instead of using stereo-pairs, structure from motion uses multiple overlapping photographs for feature extraction and 3-D reconstruction.

- Feature matching and sparse point cloud production

Key points are identified using the scale invariant feature transformation (SIFT) algorithm. The four major stages of the SIFT computation include: 1) scale-space extrema detection, 2) keypoint localization, 3) orientation assignment, and 4) key point descriptors (Lowe 2004). A 3D model and point cloud of an object is computed by using an iterative process called random sampling consensus (RANSAC) on the image dataset (Bolles & Fischler 1987). RANSAC is further used to calculate and optimise camera positions as no prior knowledge regarding exterior orientation is needed.

- Point cloud densification and surface modelling

A densification method is used to add detail to the sparse point cloud by increasing its density to create more meaningful renderings of the target object (Fukurwa 2011). Densification may exclude artefacts that otherwise may have been included in the model. Multiview-stereo (MVS) was used to aid surface rendering. Part of MVS's success is the derivation of point clouds in

weakly supported surfaces with high degree of specular reflection, transparency or lack of texture (Fukurwa 2011).

- Orthorectification

Geo-referencing using GCPs are used to transform surface models from arbitrary 3D space coordinates to real word coordinates. Geo-referencing can be undertaken at varying stages of image processing. In the context of this study, geo-referencing was performed after the final model had been computed. This three step process resulted in a DSM that was subsequently subtracted from the SUDEM (DTM) to create an nDSM.

3.3.4 Accuracy assessments

Three accuracy assessments were carried out to evaluate the quality of the derived nDSM from the different datasets. The reference extents were created by manually digitizing dwellings extents and compared to the extracted nDSM dwelling extents.

3.3.4.1 Area-based accuracy

The evaluation of area-based accuracy metrics uses a set of statistical parameters defined by McGlone & Shufelt (1994). It involves the spatial comparison of a referenced area against an extracted area using metrics such as completeness and correctness. This accuracy assessment method has been successfully employed in several studies, for example Olsen (2004), Hofmann et al. (2001), Hermosilla et al. (2011), and Champion et al. (2013). For this study, the extents of the field verified and digitized referenced dwellings were compared with extracted nDSM dwelling extents. A sample size of 50 was selected for both test site 1 and 2.

Area-based accuracy assessments are explained in Figure 3.6. If ABCD is the reference area and MNOP is the extracted area, then the following can be calculated. True positive (TP): the area present in both reference and extracted area, represented in green. False positive (FP): the area present in the reference but not extracted, represented in blue. False negative (FN): the area present in in the extracted area but not in the referenced area, represented in red. With these three types of areas defined, the following quality metrics can be defined: Completeness, which refers to the omitted area of the extracted dwelling in relation to the reference area, whereas correctness is a measure of the detected area of the extracted informal dwelling relative to the reference area (Sebari & He 2013).

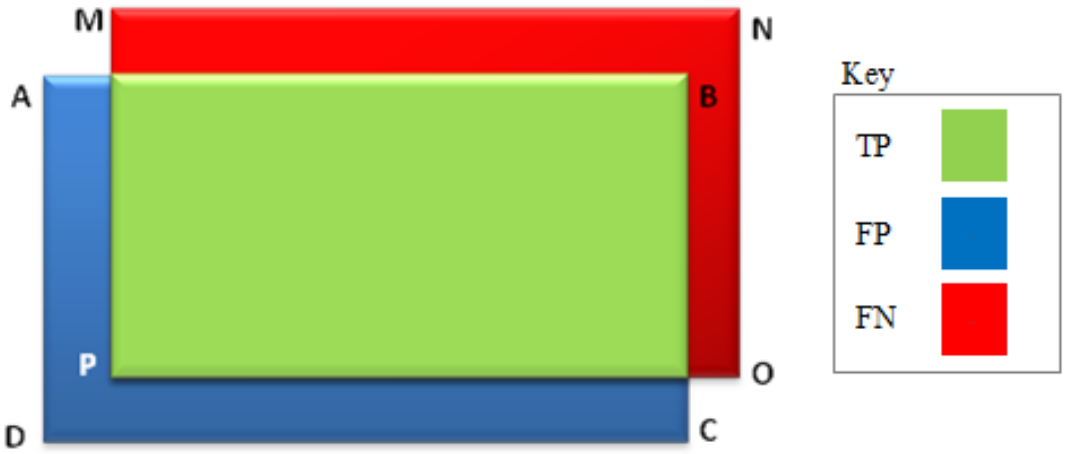


Figure 3.6 Area-based accuracy assessment method to calculate FP, FN, and TP.

Completeness and Correctness can be defined as:

$$\text{Completeness} = \frac{TP}{TP+FN} \qquad \text{Equation 3.1}$$

$$\text{Correctness} = \frac{TP}{TP+FP} \qquad \text{Equation 3.2}$$

The extracted building area provides an account of the general performance of the derived nDSM with regards to how well it represents the reference area, true positives (TP), false positive (FP) and false negatives (FN), and gives a better assessment of what extent the buildings were correctly and incorrectly detected. TP should be as close to the reference and extracted areas as possible with low FP and FN, indicating low omission and commission. Hermosilla (2011) describes that high false negatives (FN) area degrade completeness whereas high false positives (FP) area degrade correctness. Therefore a lower FP and FN value is desirable and is representative of high accuracy.

3.3.4.2 Position-based accuracy

The nDSMs were also evaluated by using a confusion matrix (Kohavi & Provost 1998). A confusion matrix shows the actual and predicted classifications performed by the classification system. For this study a binary assessment was carried out between dwellings and non-dwellings to assess the ability of the nDSMs to separate objects above the DTM. Additionally, a KHAT statistic (Congalton & Green 2009) was calculated which tests whether the values in the confusion matrix are due to true agreement, or are due to chance agreement. The sample size of 50 was used for both test site 1 and test site 2.

3.3.4.3 Vertical profiles

The vertical accuracy of the extracted nDSM was assessed using vertical profiles. This particular method was chosen over a more traditional Root Mean Squared Error (RMSE) calculation as the mean height of individual dwellings were calculated and compared to a mean field reference height and not a singular reference point. Vertical profiles of the individual dwellings were evaluated by using the field verified and digitized dwellings extents. The extents were clipped from the respective derived nDSMs and, using zonal statistics, the mean height within each reference dwelling extent were calculated (Stal et al. 2013). A sample size of 25 was selected for both test site 1 and test site 2 and reference field height measurements were recorded per dwelling.

3.4 Results

3.4.1 Area-based accuracy

The results of the area-based accuracy assessment are shown in Figure 3.7. In test site1 a total reference area of 1222.1 m^2 was calculated from the field verified and digitized dwelling extents. The highest area of extracted building was achieved by IM (1677 m^2), followed by SfM (1499.2 m^2), and LiDAR (1076 m^2). IM achieved the highest TP (1136.7 m^2) followed by SfM (1099.9 m^2), and LiDAR (912.4 m^2). Although the highest TP was achieved by IM, this does necessarily mean that IM will produce the best results, as the remaining extracted area needs to be calculated in terms of FP and FN.

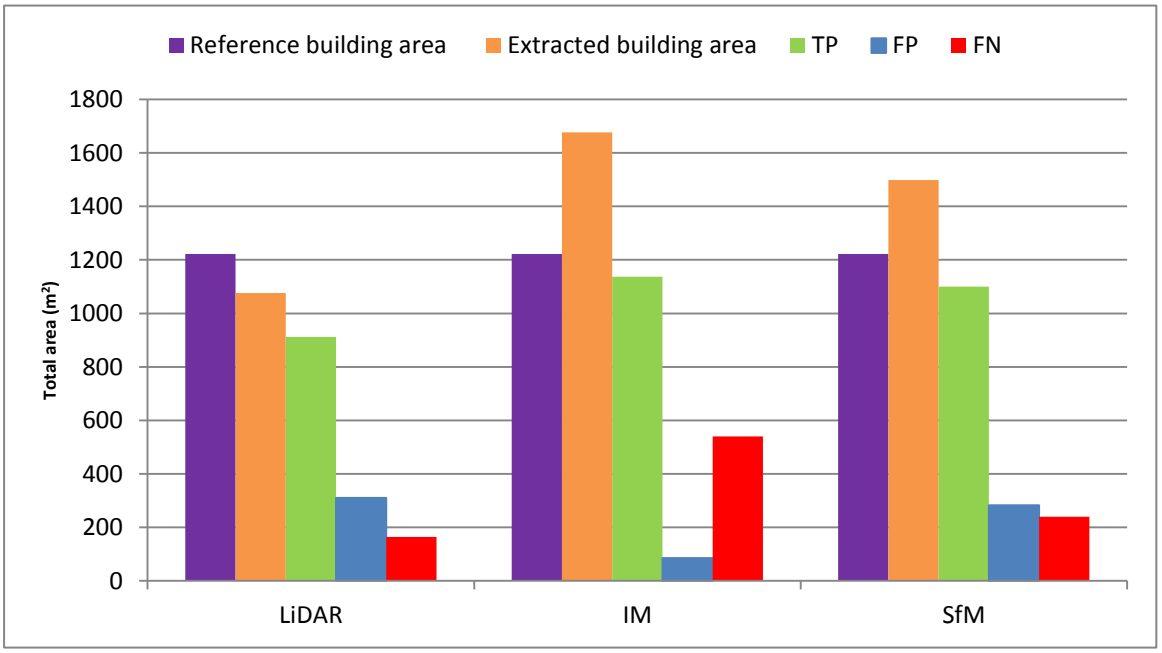


Figure 3.7 Area-based accuracy assessment for test site 1 based on 50 dwellings.

The lowest FP area was achieved by IM (85.3 m^2), followed by SfM (283.1 m^2), and LiDAR (309.6 m^2). However, the highest FN area was achieved by IM (540.2 m^2); with SfM and LiDAR achieving FNs of 239.8 m^2 and 163.6 m^2 respectively. Although IM achieved the highest TP value, a high FP value was achieved which indicates that the remaining extracted building area (1677 m^2) was attributed to incorrectly detected areas. Conversely LiDAR achieved the lowest TP (912.4 m^2) in relation to the reference area (1222.1 m^2), resulting in high FP value which indicates omitted area.

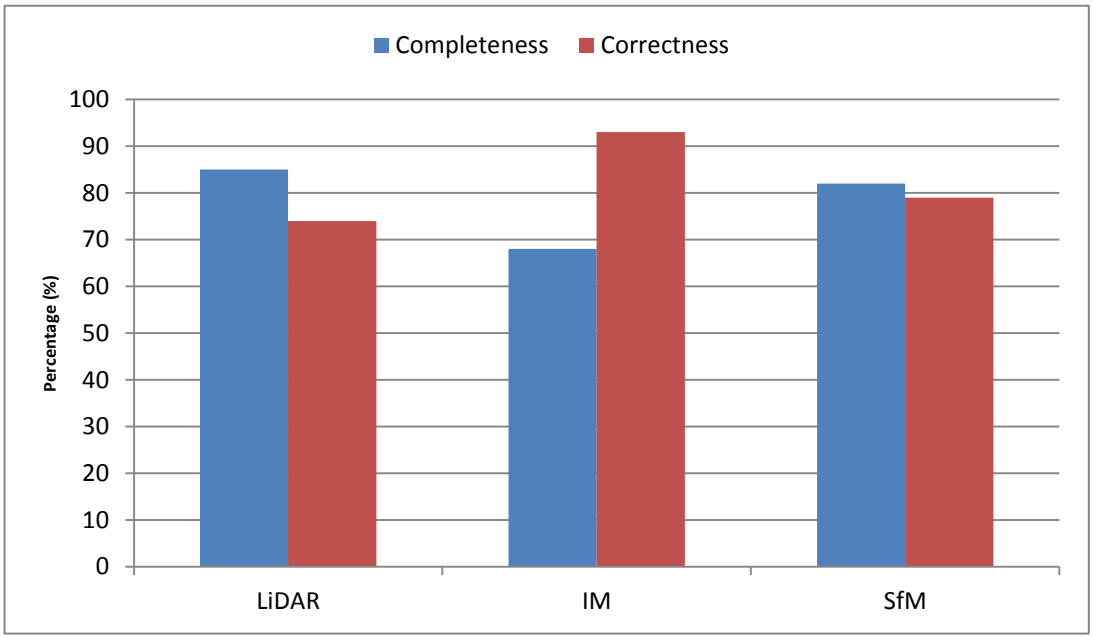


Figure 3.8 Completeness and correctness computed test site 1.

LiDAR achieved the highest completeness (85%) followed by SfM (82%), and IM (68%). The highest correctness was achieved by IM (93%), followed by SfM (79.5%), and LiDAR (74.6%). Figure 3.9 shows an example of TP, FN and FP for each of the nDSMs in test site 1. It can be seen that LiDAR displays a smaller FN area compared to SfM and IM, whereas SfM has a greater TP and FP.

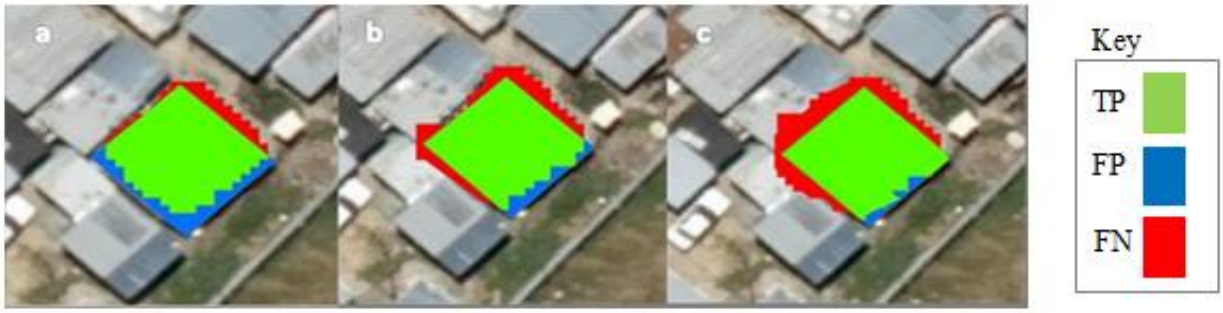


Figure 3.9 (a) LiDAR (b) IM and (c) structure from motion (SfM) extracted nDSMs are shown against reference dwellings extents in test site 1.

In site 2 a total reference area of 1214.8 m^2 was calculated from the field verified and digitized dwelling extents. The highest TP area was recorded for LiDAR (1034.4 m^2), followed by SfM (973.4 m^2), and IM (820.51 m^2). A lower FP and FN value was achieved by LiDAR (180.3 m^2 and 146.9 m^2) in comparison to SfM (241.5 m^2 and 134.1 m^2), and IM (411.7 m^2 and 199.1 m^2). LiDAR's FP was less than SfM and less than half that achieved by IM, revealing the difference in omitted area between the derived nDSMs. The similarity in extracted building area, TP and FN values between LiDAR, SfM and IM indicate similarity of included/misclassified areas. As a result the highest correctness and completeness in was achieved by LiDAR (85%), followed by SfM (80%), and IM (67%) as shown in Figure 3.11.

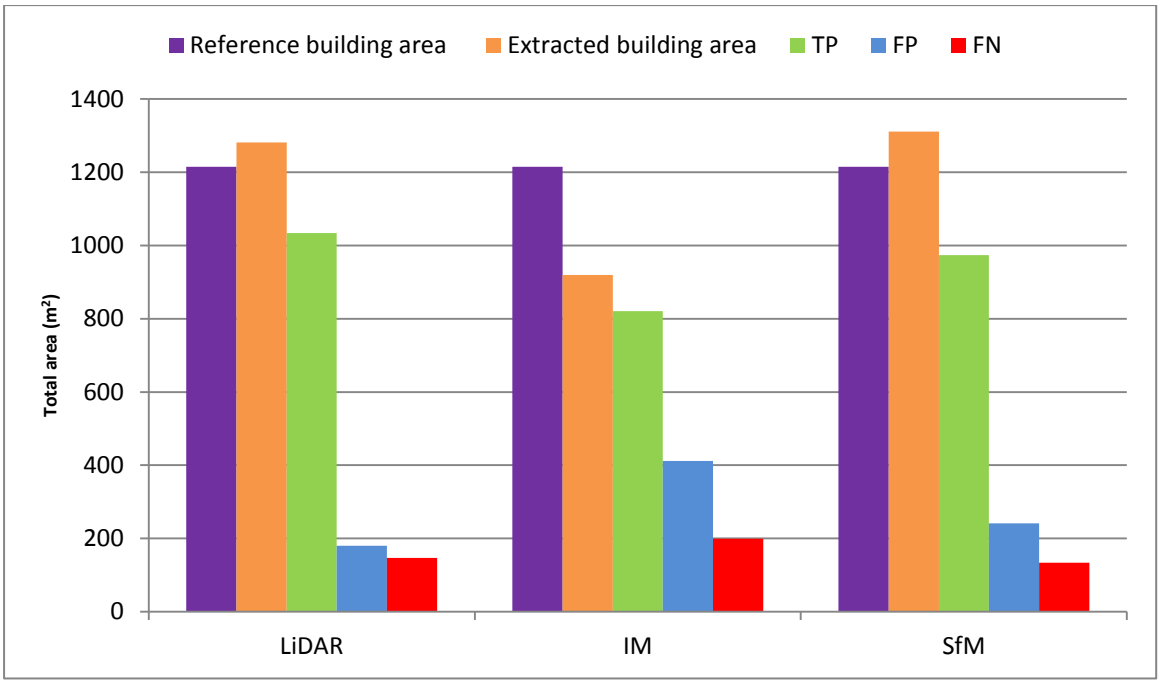


Figure 3.10 Area-based accuracy assessment for test site 2 based on 50 dwellings.

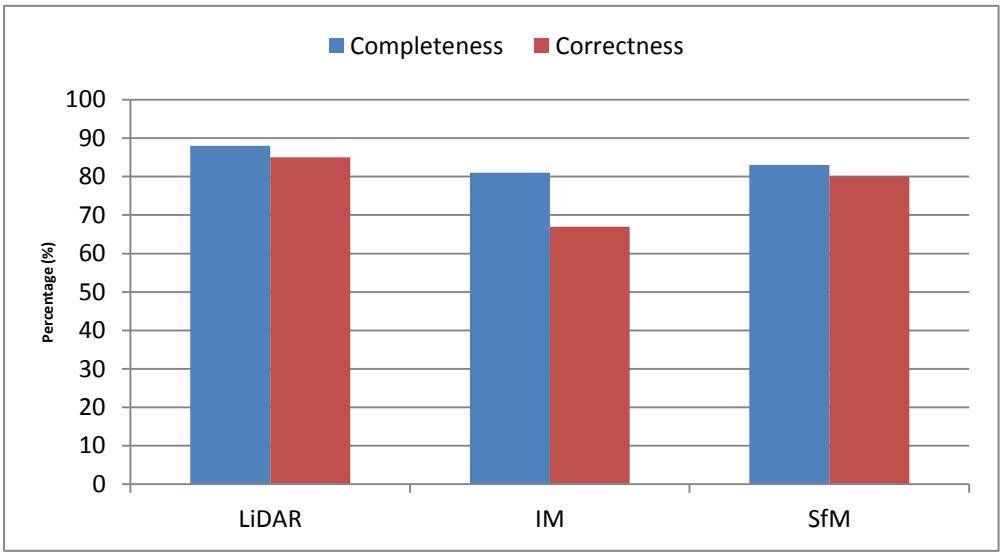


Figure 3.11 Completeness and correctness computed for test site 2.

3.4.2 Position-based accuracy

LiDAR, SfM and IM achieved varied overall accuracies and Kappa values for the two test sites. In test site 1, the highest producer's accuracy was achieved by LiDAR (100%), followed by SfM (89.6%), and IM (81%). In contrast, the highest user's accuracy was achieved by IM (96.7%) and the lowest by LiDAR (93.6%). The highest overall accuracy and KHAT was achieved by LiDAR (96.3%, 0.93), followed by Structure from Motion (91%, 0.74), and IM (86.6%, 0.81). The results in test site 1 indicate that very little omission and commission occurs between dwellings and its surroundings despite the dense dwelling arrangement. In test site 2, the highest producer's and user's accuracy was achieved by LiDAR (99.3% and 95%), followed by SfM (97.5% and 93.4%), and IM (82.1% and 87.3%). The highest overall accuracy was achieved by LiDAR with 97.1% (KHAT: 0.90) followed by SfM with 93.4% (KHAT: 0.81), and IM with 87.3% (KHAT: 0.73).

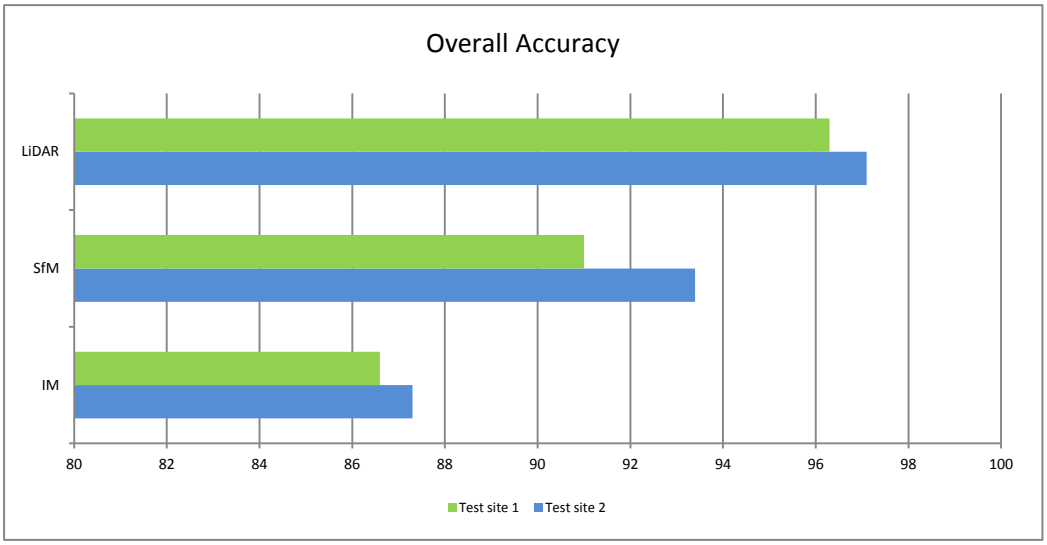


Figure 3.12 Overall accuracy percentages achieved for LiDAR, SfM and IM in test site 1 and 2.

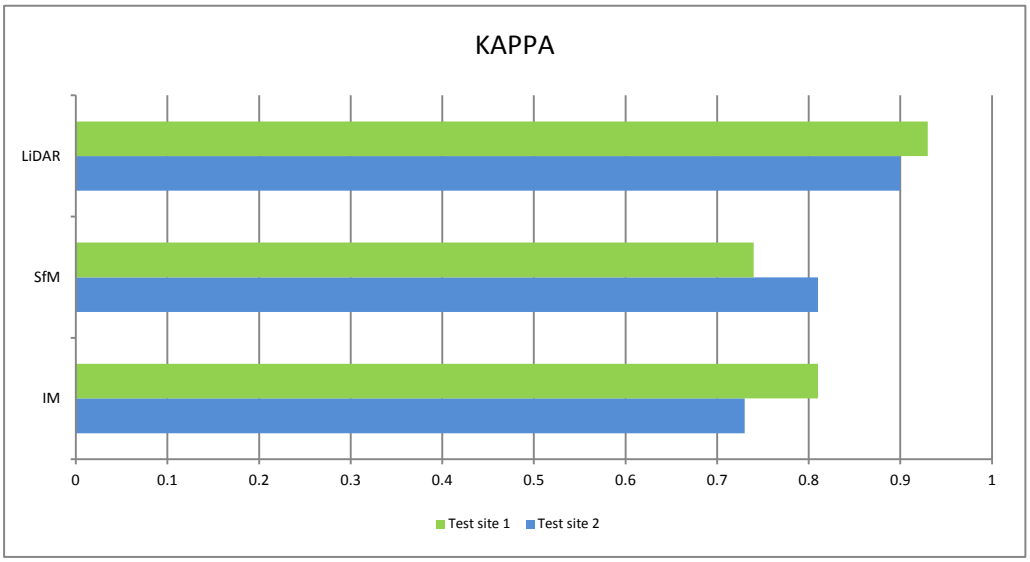


Figure 3.13 Kappa values achieved for LiDAR, SfM and IM in test site 1 and 2.

3.4.3 Vertical profiles

In test site 1, the average reference height recorded was 2.12 m. The most accurate mean height calculated was using LiDAR (2.9 m), and the least accurate was IM (3.14 m). In test site 2 the reference height recorded was 1.82 m. The similarities between the mean heights can be noted with LiDAR (2.08 m), IM (2.10 m), and SfM (2.14 m) having an average separation distance of 6 cm, whereas in test site 1 an average separation distance of 75cm can be observed as shown in Figure 3.14.

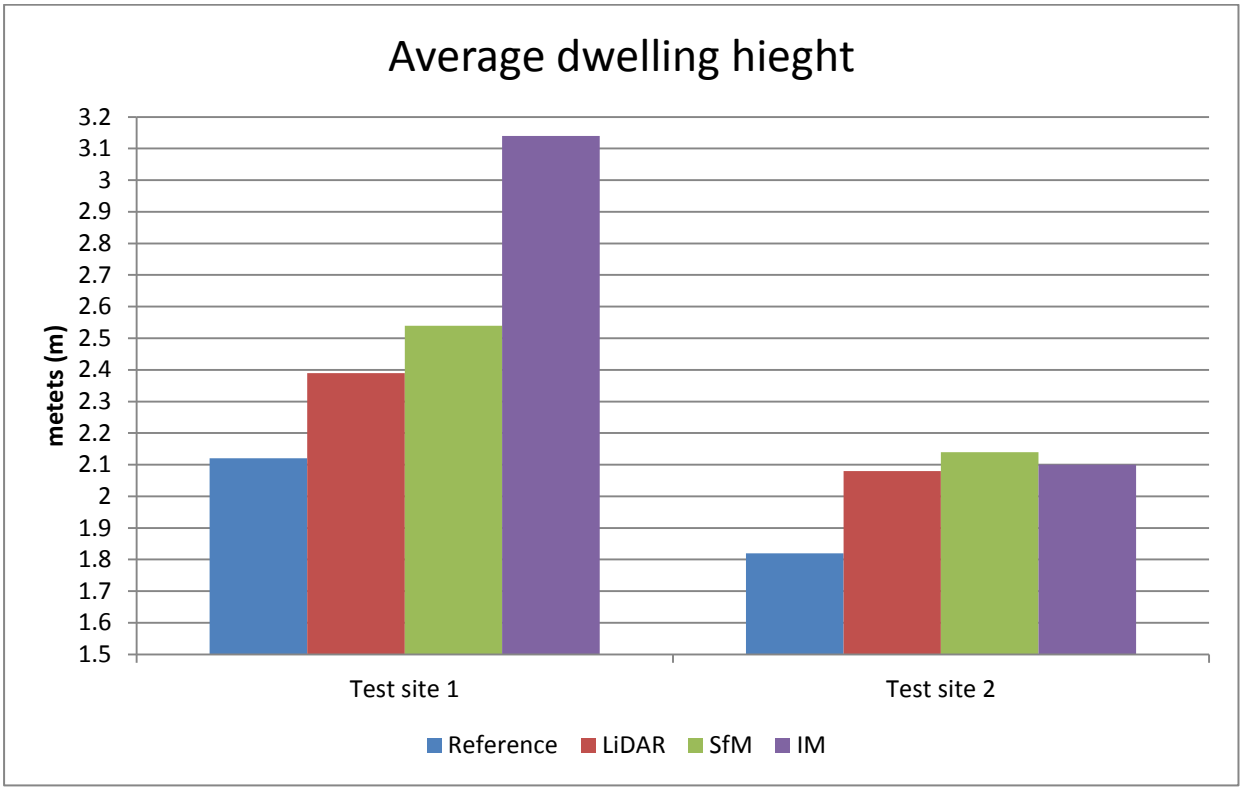


Figure 3.14 The average height of individual dwellings in test sites 1 and 2 for LiDAR, SfM and IM.

The vertical profiles for 25 dwellings for test site 1 are shown in Figure 3.15. It can be noted that the trend lines of LiDAR, IM and SfM are notably different. For example a significant difference at dwelling number 8 can be seen between LiDAR (1.8 m), IM (3.1 m) and SfM (2.6 m). Likewise at dwelling number 15 between LiDAR (2.0 m), IM (3.4 m), and SfM (2.9 m). Overall it can be seen that LiDARs trend line more closely represents the reference trend line when compared to IM and SfM.

In test site 2 (Figure 3.16), the vertical profiles trend lines of LiDAR (2.08m) more closely match the reference (1.83m) than SfM (2.10m) or IM (2.14m). Individual dwelling heights are very similar between LiDAR, SfM and IM at dwelling number 3 (2.23 m, 2.56 m, 2.25 m), dwelling 12 (2.16 m, 2.47 m, 2.13 m), and dwelling 22 (1.89 m, 1.87 m, 1.98 m).

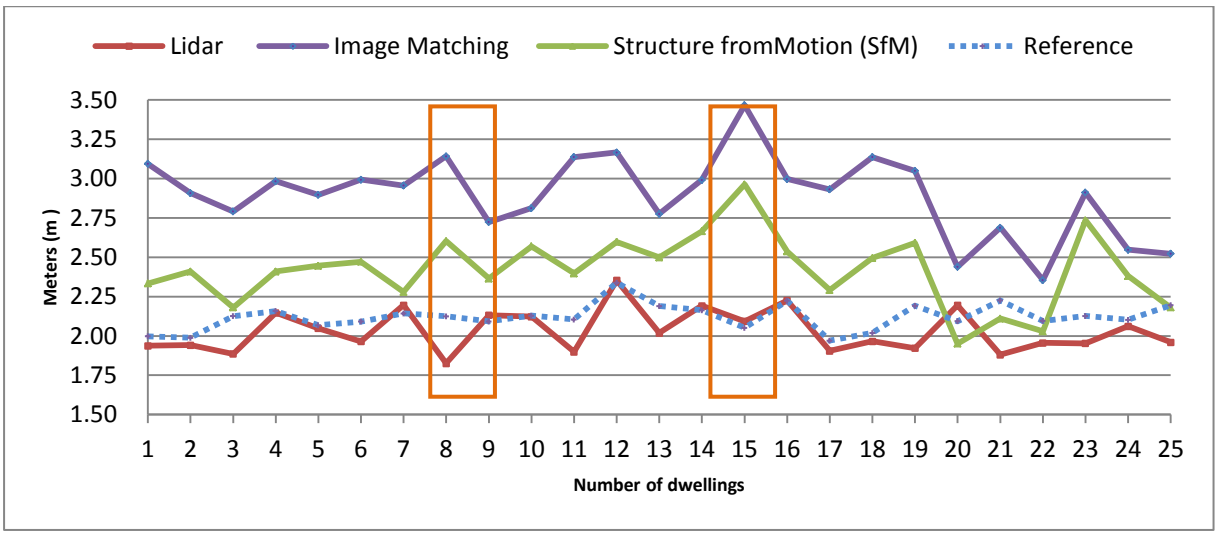


Figure 3.15 Vertical profiles of 25 dwellings in test site 1 for LiDAR, SfM and image matching.

The effect of the spatial structural arrangement of the informal settlements can clearly be seen when comparing the vertical profiles trend lines of LiDAR, SfM and IM of test site 1 with test site 2. Although LiDAR, SfM and IM follows each other's trend lines well, overall it is does not closely represent the reference trend line, such as LiDAR in test site 1. Overall LiDAR yielded the highest accuracy for both test sites with the closest average dwelling height and the individual dwelling height to the reference.

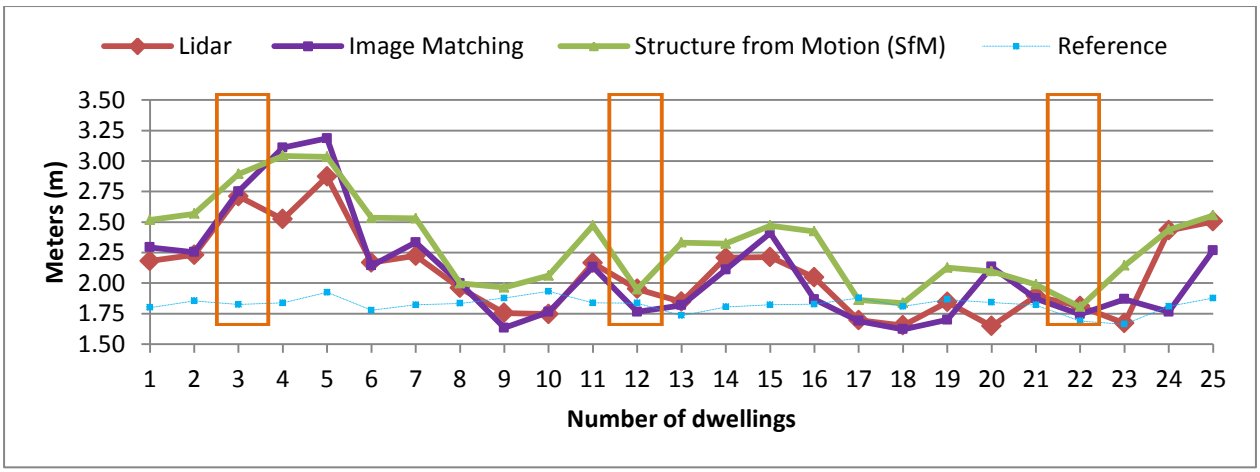


Figure 3.16 Vertical profiles of 25 dwellings in test site 2 for LiDAR, SfM and image matching.

3.5 Discussion

In this study high resolution nDSMs from aerial imagery using SfM and IM, and LiDAR data was derived for informal dwellings. Previous studies have assessed IM and LiDAR (Moreira et al. 2013), IM and SFM (Andrews et al. 2013) or LiDAR and SfM (Maiellaro, Zonno & Lavalle 2015). This study is a first attempt at evaluating the all these above methods together. It further evaluated nDSM accuracy using vertical profiles, area-based and positional-based accuracy assessment metrics. The results are significant within a South African and global context, given the increased demand for high resolution digital elevation models that can be used to fulfil the geospatial information required for the planning and modelling of informal settlements.

3.5.1 Area-based accuracy

It can be observed that the test site 1 results are respectively different to that of test site 2, and are related to the density of dwellings within the informal settlements. Overall a higher completeness can be noted in test site 2 when compared to test site 1 and is linked to the decrease in FP areas in relation to FN areas between the two informal settlements. In both test sites 1 and 2, high completeness was achieved (< 80%) by LiDAR and SfM which shows that a small proportion of misclassified area (FN) was included in relation to the reference dwelling area. A lower completeness achieved by IM in test site 1 (68%) however suggests that high percentage of dwellings were misclassified when compared to test site 2 (81%). In contrast IM achieved the highest correctness (93%) in test site 1 and the lowest correctness in test site 2 (67%). The reason for the contrasting results is due to IMs over representation of dwelling extents with low omission (FP) and high included areas (FN) in test site 1 and vice versa in test site 2. To better understand the relationship of completeness and correctness, different combinations of FP and FN are shown below in Figure 3.17 as high FN area was found to degrade completeness and

Figure 3.17 a) no FN and high FP, b) high FN and no FP, c) high FN and high FP, and d) no FN and No FP

high FP area was found to degrades correctness (Hermosilla 2011). Therefore a lower FP and FN value is desirable and is representative of high model accuracy.

The results of this study is in agreement with Awrangeb and Fraser (2013) who found that completeness was lower in a dense building environment consisting of complex shapes (83.8%) with more reflective surfaces and its corresponding albedos (Stal et al. 2013, Böhler, Bordas, and Marbs 2003) when compared to an area with 'detached houses with many surrounding trees' (85.7%). In addition the results achieved in this study compare favourable to Joshi (2010) who extracted completeness and correctness in sparse (93% and 92%) and dense building areas (85% and 86%). The authors found that irregular building dimensions, high building density, and buildings with similar spectral characteristics contributed to a decrease in completeness and correctness. These contributing factors, characteristically describe test site 1, and from this study it can be noted that opposite holds true, that low building density and buildings with dissimilar spectral characteristics contributed to the increase in completeness and correctness as observed in test site 2. Thus the higher accuracy achieved in test site 2 demonstrates the effect of informal settlement spatial structure on the performance of area-based metrics (Joshi 2010).

3.5.2 Position–based accuracy

No significant differences in a position-based accuracy were found in test site 1 when compared to test site 2, as overall accuracies were similar for LiDAR (96.33% and 97.17 %), IM (86.65% and 87.38%), and Structure from Motion (91.00% and 93.46%). The accuracies achieved can be translated to the errors of omission and commission to better understand how error is represented in the nDSMS. LiDAR achieved low errors of omission and commission in both test site 1 and 2, whereas a higher error of commission was achieved for IM in both test site 1 and 2. In contrast, SfMs achieved both high errors of commission in test site 1 and omitted area in test site 2. In the context of this study, the results is significant as it provides an understanding on how well the nDSMs represent the dwellings above the DSM, especially as it will be used as an ancillary dataset for classification and should represent minimal non-dwellings objects. Thus it is important to select a digital model that preserves dwelling boundaries, with low errors of commission and omission to avoid misclassification. Overall the LiDAR results demonstrate the best ability to meet these criteria. The results achieved in this study compare favourably to Ekhtari et al. (2008) who detected buildings using LiDAR with an overall accuracy of 95.1%. Nex, Dalla Mura and Remondino (2013) detected dense buildings using IM and achieved users accuracy, producers and overall accuracy (85.7%, 97.4% and 87.0%). They found that lower building extraction success was due to taller buildings partially occluding or casting shadows

over smaller buildings. The high results achieved in this study demonstrate that these aspects did not affect nDSM extraction due to the relatively low height informal dwellings.

3.5.3 Vertical profile

One of the main concerns in urban modelling is evaluating vertical accuracy, as generated models often do not exactly match the shape (z-axis) of the corresponding buildings (Poon et al. 2007). In test site 1 the vertical profiles did not closely represent the average reference height of 2.12 m, as the mean heights for individual dwellings are significantly different for SfM, LiDAR and IM. In contrast, the Joosdale informal settlements vertical profiles do closely represent the average reference height of 1.82 m. It can be noted that IM and SfM vertical profiles follow LiDAR profiles better in open areas (test site 2) than dense areas (Agugiaro, Poli and Remondino 2012), and that LiDAR more closely represents building reference profiles in test site 1 and 2. The difference in vertical profiles when compared to the reference height in the two informal settlement suggest that an increased accuracy can be achieved in a sparse building setting, and that vertical accuracy is adversely effected in a dense building setting (Joshi 2010).

3.5.4 Accuracy versus model limitations

The selection of a preferred data source for digital elevation modelling is finding a balance between many factors including the project constrains such as cost involved in acquisition and processing, as well as inherent limitations and challenges of a particular DEM extraction method (Martina & Ginzler 2012; Remondino et al. 2014). It is important to be mindful of limitations in model accuracy as errors will propagate through to model predictions (Fryer et al. 1994). It was noted that substantial uncertainty exists with image acquisition and processing for IM, SfM and LiDAR datasets due to the structural complexity, diversity and high degree of detail within informal settlements (Stal et al. 2013).

3.5.4.1 LiDAR

It was observed that LiDAR systems are not influenced by shadows, they generally preserves surface discontinuities and allow for denser and more accurate measurements to be performed (Remondino et al. 2013). However the quality of LiDAR measurements suffers heterogeneously from different type of reflective surfaces that is present in an informal settlement (Stal et al. 2013). This is caused by a variety of materials used in the construction of dwellings and its corresponding albedos (Böhler, Bordas, & Marbs 2003) as is evident in dense Danoon informal settlement. The accuracy and performance of LiDAR can be reduced by model noise (Schaer et al.2007), low spatial detail, and secondary reflections close to vertical structures (Stal et al. 2013). This can result in indistinct dwelling boundaries that mainly occur when there are less

than one metre gaps separating dwellings such as in this study. Similarly vegetation, specifically tall trees adjacent to dwellings, can cause interference during LiDAR acquisition, as tall trees can be erroneously included as buildings (Cheuk & Yuan 2009). The acquisition time and cost of LiDAR is relatively cumbersome and more expensive than systematically acquired stereoscopic images (Gehrke et al. 2010; Remondino et al. 2013) as LiDAR acquisition have namely longer flight times, narrow swaths and a more complex flight plans (Baltsavias & Gruen 2003). However, the computational cost of LIDAR is lower when compared with SfM and IM, and allows higher automation of data processing and relatively short production times (Stal et al. 2013).

3.5.4.2 Image Matching and Structure from Motion

The significant technological improvements have seen photogrammetry based algorithms re-emerge as a competitive technology to range finding sensors such as LiDAR, which has grown in popularity as a fundamental source of dense points cloud mapping and visualization (Remondino et al. 2013). In this study the main limitations of IM were 1) low texture or repetitive texture on dwelling roofes, 2) radiometric artefacts caused by specular reflectors such as reflective metal material used in construction of the dwellings, 3) occlusion and shadows as a result of taller dwellings adjacent to each other, and 4) trees adjacent to dwellings (Gruen 2012; Zeng et al. 2014). These artefacts can result in a noisy point cloud and/or difficulties with feature extraction (Remondino et al. 2013). Although mage matching is a more labour intensive task, unlike LiDAR and SfM systematic error corrections and error metrics can be explicitly stated during the calibration process (Remondino et al. 2014). The progress, prospects, development and status of IM in photogrammetry can be found in Gruen (2012).

Unlike IM, SfM eliminates the requirement for manual identification of orientation parameters by supporting changes in camera perspective and image scale (Westoby et al. 2012). This flexible, user-friendly approach represents a major advancement in the field of photogrammetry as data acquisition is not restricted to trained professionals or researchers, but can be acquired with a point-and-shoot digital camera (Westoby et al. 2012). However, this relatively inexpensive and low cost approach is characterized by lengthy processing time and significant post-processing load that is largely dependent on the complexity of image texture and size of the image photoset (Westoby et al. 2012). SfM is also classified as a ‘black-box’ tool where bundle adjustment divergence or geometric deformations may occur as there is no predictable result at the end of the process (Remondino et al. 2014). Remondino et al. (2014) provide a list of requirements to assess if feature extraction and reconstruction algorithms fulfil the prerequisites

to be successful, powerful and reliable. In the context of this study dense point clouds with a sufficient resolution to describe the objects in regions with poor texture or illumination and scale variations were needed.

3.5.5 Conclusion

This paper represents a first attempt to simultaneously compare and evaluate nDSMs extracted from high resolution aerial photographs using IM and SfM to LiDAR data. The accuracy of the respective nDSMs was evaluated using vertical profiles, area-based, as well as positional-based accuracy assessment metrics. The results demonstrated that these technologies and techniques can be used to extract nDSMs within informal settlements. It was interesting to note how accuracies varied between the three accuracy assessments depending on the structural arrangements of dwellings in the respective informal settlement. A general increase in accuracy was observed/ recorded in the sparser and open informal settlements, and is directly related to limiting factors in technique and data or the method of nDSM extraction. LiDAR achieved the best results overall in all three accuracy assessments, however SfM was found to be a suitable/ potential alternative to the more expensive LiDAR and IM approaches, especially considering improvements in point-and-shoot cameras and computing power. An important consideration before selecting a preferred dataset or method to map informal settlements is the balance between the desired output accuracy of the nDSMs, the cost involved in acquisition, the time spent during processing, and project constraints and requirements.

CHAPTER 4: IDENTIFICATION OF INFORMAL DWELLINGS FROM HIGH RESOLUTION IMAGERY USING AN OBJECT-BASED IMAGE ANALYSIS APPROACH*

4.1 Abstract

This study tested the utility of high resolution WV-2 and aerial imagery within an OBIA environment for the objective identification of image segmentation, feature selection and classification parameters in the context of informal settlement mapping in the City of Cape Town. The supervised segmentation parameter tuner (SPT) tool was used to derive optimal segmentation parameters, and was evaluated using an area-based accuracy assessment which resulted in high compactness and correctness. To reduce data dimensionality and optimize the classification process, the RF and CART algorithms was used to select WV-2 and aerial imagery features. The RF algorithm reduced the original WV-2 and aerial imagery features by 23% and 53%, whereas the CART algorithm reduced the same feature set by 95% and 91% respectively. Nevertheless although different feature subsets were selected by the RF and CART algorithm for the WV-2 and aerial imagery, similar classification accuracies were achieved in all the test sites. In addition this study is the first to include an nDSM to assist classification of informal dwellings within an informal settlement.

4.2 Introduction

The rapid migration of people to urban areas has led to urban expansion in the form of peripherization - the development of informal settlements or slums in peri-urban areas, which fail to comply with the land administration protocols (UN-Habitat 2012/2013). The UN-Habitat (2012/2013) estimated that 863 million people worldwide are living in informal settlements or slum conditions in contrast to 760 million in 2000 and 650 million in 1990. Sub-Saharan Africa recorded the highest informal settlement prevalence in world with 62%, approximately 199.5 million people living in informal settlement conditions in urban areas. This estimate is more than the 190.7 million people in Southern Asia (35%), 110.7 million people in Latin America and the Caribbean (23.5%) and 11.8 million people in Northern Africa (13.3%). In 2001 approximately

---

*This chapter will be submitted for publication and will conform to the prescribed structure of that journal. Some of text and figures used in this chapter are identical to those of previous chapters, as the same data and methods were used.

16.7% of South Africa's population lived in informal settlements, this figure rising to 18.2% in 2011 (Census 2011). The dynamic spatial-temporal behaviour, the relatively high inner-structural heterogeneity (Hoffman et al. 2008; Pinho et al. 2011; Shekhar 2012) and microstructure of informal settlements do not allow generically applicable mapping solutions to be produced (Taubenbock & Kraff 2013). This is demonstrated by the lack of alignment regarding data collection methodologies and census inconsistencies at municipal, provincial and national scale (Housing Development Agency 2012). Current estimates are based on the collective number of informal settlements and the total area covered in square kilometres but gives no indication of individual dwelling counts within informal settlements; as informal settlements are often omitted from formal statistical assessments (Kohli et al. 2013). A suitable intervention effort is thus required for methodology development that is both robust and transferable. A robust and transferable methodology can provide data with consistent and predictable results that can assist in the evaluation of performance and intervention based programs (Kohli et al. 2012). This can facilitate programs in South Africa such as the 're-blocking' policy that the City of Cape Town has recently adopted. Re-blocking involves reconfiguration and repositioning of dwellings within informal settlements to improve the utilization of space for planning of provisions, installation of local government services, and for disaster risk mitigation (South African SDI Alliance 2013). By mapping informal settlements it is possible to identify preferred locations allowing upgrading efforts to be concentrated on established settlements, where the probability of achieving successful re-arrangement is higher (Informal Settlements Handbook 2009).

At the expert group meeting on slum identification and mapping (Sluizas et al. 2008), it was concluded that remote sensing-based approaches and VHR imagery provide a flexible and useful method to address the complex morphology of informal settlements (Hofmann et al. 2008; Salehi et al. 2011; Kit et al. 2012 ; Shekhar 2012; Kohli et al. 2013). One of the several approaches that have the potential to address the relatively complex and undefined urban morphology of informal settlements include object-based image analysis (OBIA) (Hofmann et al. 2008 ; Salehi et al. 2011; Shekhar 2012; Kit et al. 2013; Kohli et al. 2013). OBIA is a semi-automated approach that attempts to overcome the limited spectral information of pixels by combining spectral, spatial, and textural information to form homogenous image objects through segmentation (Baatz and Schäpe 2000; Blaschke 2010). The image objects are subsequently used as input for image feature selection that facilitates rule-set development or with an application of a classification algorithm (Benz et al. 2004).

The relatively low number of international publications on OBIA-based informal settlement mapping confirms the challenges for accurate methodology development, given the diverse definition, geographical setting and local conditions of informal settlements (Kit 2012; Kohli et al. 2012). The few studies that have demonstrated the use of OBIA for classification have integrated the relative knowledge of the real world characteristics of informal settlements, such as the spectral, the geometric, contextual properties, and relationships between objects (Hoffman 2001; Hoffman et al. 2008; Shekhar 2008; Kohli et al. 2012; Kohli et al. 2013). A pioneering study by Hoffman (2001) used OBIA to identify informal settlements from IKONOS imagery in the City of Cape Town. Informal settlement classification was undertaken using sub-classes that described settlement forms (dense, medium, new and bright) based on complex hierarchy and class descriptions such as textural and spectral features. The author found that the ability to detect informal settlements was dependent on the spatial resolution of the imagery. No quantitative results were presented as findings of the study. This research was later improved by Hofmann et al. (2008) who showed that several modifications were required when applying extraction methods to a QuickBird scene in Brazil. The modifications included simplified and pruned class-hierarchies making the chosen class descriptors in theory more transferable to comparable scenes. The results of this study demonstrated that the selection of a strategy for informal settlement segmentation and classification is data and context specific. Tiede et al. (2010) extracted structures in refugee and IDP (internally displaced persons) camps in West Darfur using GeoEye-1 imagery. Classification was performed by incorporating dwelling spatial characteristics and limited use of spectral threshold values. The developed rule set was transferred to secondary scenes with minimal changes to the master rule set. Visual interpretation was used to validate results. The results showed high agreement of absolute numbers for automatically extracted dwellings (15 349) versus visually extracted dwellings (14 261). Shekhar (2012) delineated informal settlements in Pune City, India using QuickBird imagery. The study highlighted the efficacy of the developed methodology to discriminate informal dwellings by describing typical characteristics of these settlements. Fuzzy membership functions of texture, geometry and contextual information was used to achieve an overall accuracy of more than 87%.

Kohli et al. (2011) expanded upon the work of Hofmann et al. (2008) and developed a generic slum ontology (GSO), which was used as part of a conceptual classification OBIA schema. Kohli et al. (2012) followed an ontological approach to conceptualise informal settlements using class indicators of the built environment in Ahmedabad, India. Kohli et al. (2013) used textural features such as entropy and contrast derived from a grey level co-occurrence matrix (GLCM) combined with an adapted Global Slum Ontology (GSO) to identify informal settlements in

Ahmedabad, India using GeoEye-1 imagery. The OBIA-based classification was applied on three different subsets with minimal adaption and achieved final accuracies ranging from 47% to 68% respectively. The results showed that visually different urban patterns could be classified using a combination of different texture features to increase classification accuracy. Despite the available literature, no studies have tested the objective identification for defining image segmentation, feature selection and classification parameters together in the context of informal settlement mapping. An objective identification approach is important, as the selection of parameters is usually user-defined, error-prone, and the interpretation is biased by human subjectivity (Baatz and Schäpe 2000; Benz et al. 2004; Hofmann, Strobl & Blaschke 2008; Belgiu et al. 2011; Arvor et al. 2013). To reduce the error associated with human subjectivity, several different methods have been proposed for the objective identification of both segmentation parameters and feature selection methods for classification.

It is well understood that accuracy and reliability of classification largely depends on the accuracy of the segmentation method and strategy (Baatz and Schäpe 2000; Benz et al. 2004). Several segmentation approaches have been developed and tested such as by Baatz and Schäpe (2000), Feitosa et al. (2008), Zhang et al. (2008), Martha et al. (2011), Johnson and Xie (2011) and Drăguţ and Eisank (2011). For example Drăguţ et al. (2010) created a generic automated segmentation tool that detects patterns in data called Estimation of Scale Parameters (ESP). ESP automatically identifies patterns in the data at three different scales, from finer to larger objects in a data-driven approach (Belgiu et al. 2011). Furthermore, the number of available object features for classification (spatial, spectral, contextual, etc.) makes a detailed quantitative exploratory analysis of every individual image feature extremely time-and-effort intensive, and has led to the introduction of feature selection methods (Novack et al. 2008; Laliberte, Browning & Rango 2012). Feature selection methods are aimed to improve classification accuracy by selecting an optimal subset of features in which redundant and irrelevant features have been removed (Hapfelmeier and Ulm 2013). It is also important to identify significant features as a comprehensive feature extraction methodology is the precondition for successful work with image objects (Nussbauma et al. 2008). Various feature selection and classification methods exist including Relief-F (Novack et al. 2011), principle component analysis (Pal and Foody 2010), Jeffreys–Matusita distance (Laliberte, Browning & Rango 2012), Random Forest (Novack et al. 2011), and Classification and Regression Trees (Tayyebi and Pijanowski 2013).

The aim of this study was thus to develop a robust methodology using high resolution imagery and systematic approaches for the objective identification of segmentation, feature selection and

classification parameters for mapping dwellings within informal settlements. To achieve this objective, 1) a supervised tool for optimally deriving segmentation parameters was assessed, 2) the use of various spatial, spectral, textural features and contextual information for classification was evaluated, 3) the utility of feature selection methods to define an optimal feature subset for classification was examined, 4) several classification models using both a supervised and rule-based approach was trained and tested, and 5) the robustness of the models by application to secondary independent sites was assessed. A workflow of the methods used in this chapter is presented in Figure 4.1.

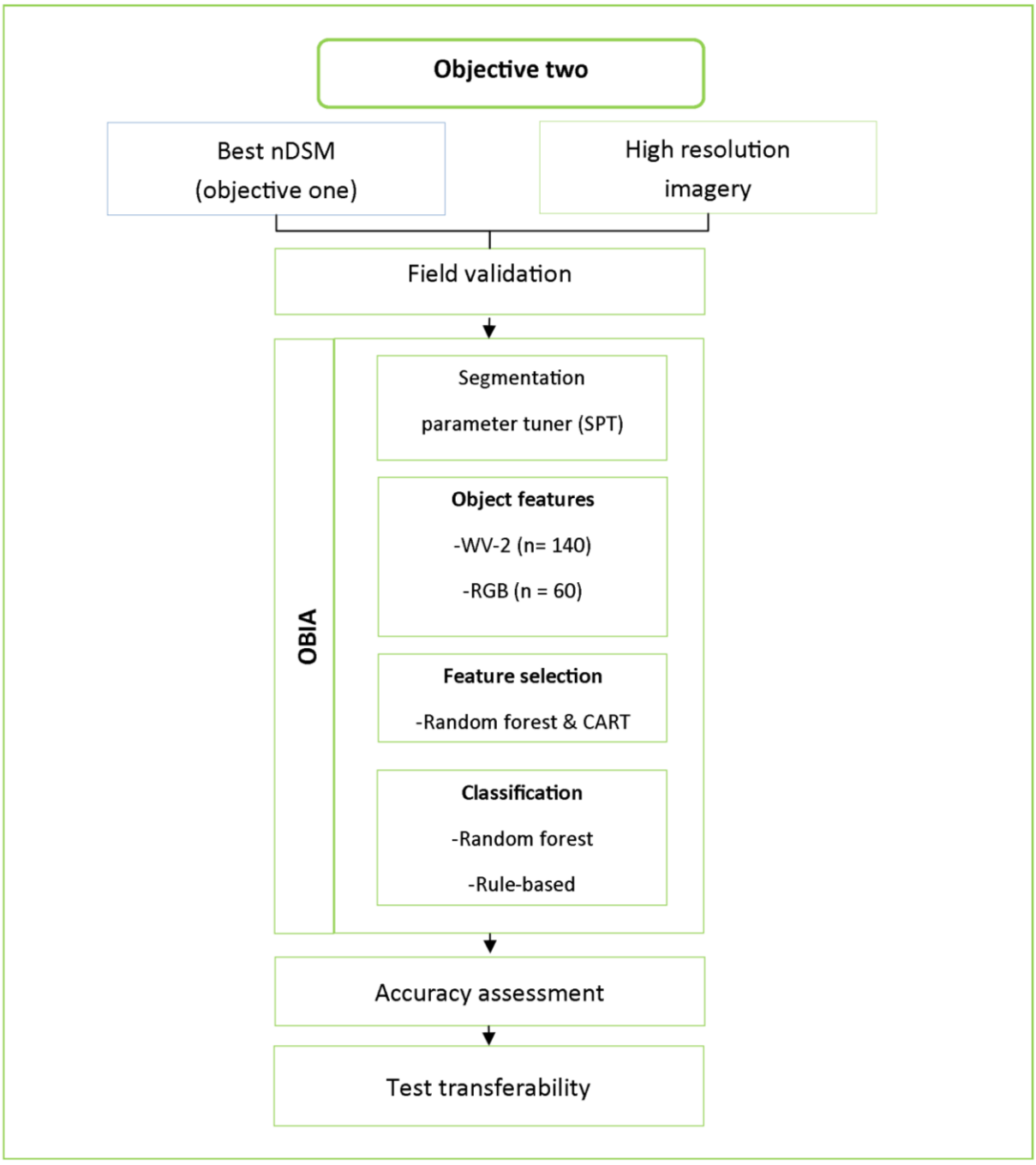


Figure 4.1 Workflow of the methods used in chapter 4. The best nDSM from chapter 3 will be used in this chapter as ancillary classification dataset.

4.3 Material and methods

4.3.1 Study sites

The study area seen in Figure 4.2 is situated in the Western Cape, South Africa. Three different types of unstructured informal settlements were chosen for this study. Study site 1 has high density, with very crowded and clustered dwellings. Study site 2 is a relatively open area with sparse dwellings and low density, and study site 3 is highly populated in an open area and medium density. The topography is all three study sites are fairly levelled. Study sites 1, 2 and 3 are consistent with properties described by Mason and Baltsavias (1997). Majority of the dwellings are single storied structures, display simple geometry (4-sided), constructed from diverse materials with variable texture and colours, and with separation distance of ~1 m between dwellings. In this study, study site 1 will be used as the training site and study site 2 and 3 will be used as test sites.

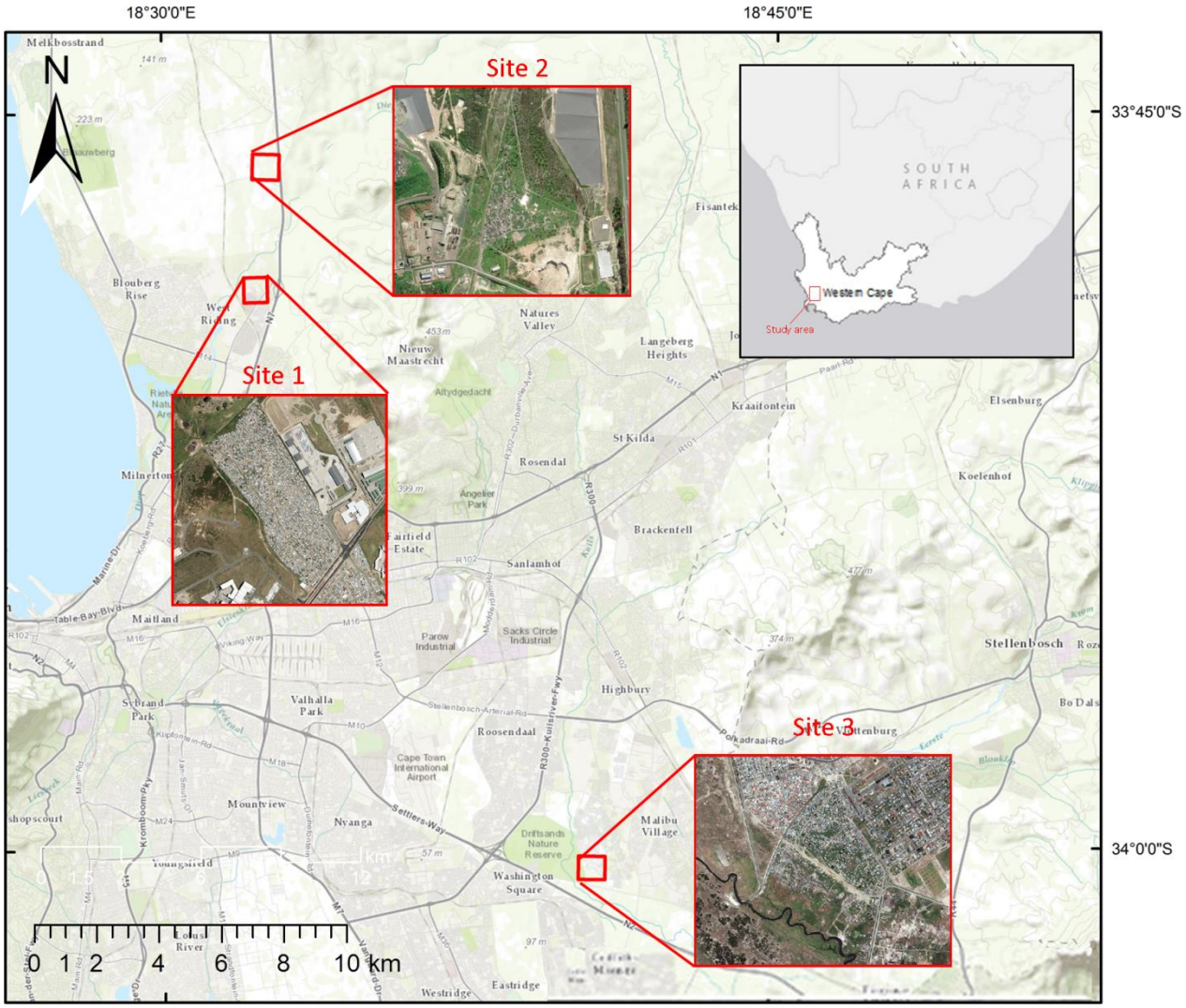


Figure 4.2 Location of the study area with WorldView-2 inset overlaid on ESRI ArcMap imagery base map.

4.3.2 Image and field data

WorldView-2 (WV-2) imagery was acquired as an 8-band map-scale orthorectified product. WV-2 comprises eight multispectral bands with a spatial resolution of 1.85 m, and a single panchromatic band with a spatial resolution of 0.46 m (DigitalGlobe 2009) (Table 4.1). WV-2 imagery was used for this study, as the additional coastal blue, yellow, red edge, and near infrared (NIR) bands could potentially improve classification results (Novack et al. 2011).

Table 4.1 Characteristics of WorldView-2 imagery.

Spectral bands	Wavelength (nm)	Spatial resolution (m)
1.Coastal blue	400–450	1.85
2.Blue	450–510	1.85
3.Green	510–580	1.85
4.Yellow	585–625	1.85
5.Red	630–690	1.85
6.Red-edge	705–745	1.85
7.Near infrared 1	770–895	1.85
8. Near infrared 2	860–1040	1.85
9.Panchromatic	450–800	0.46

Source: DigitalGlobe 2009

Very high resolution (VHR) (12.5 cm) orthorectified digital aerial imagery for test site 1, 2, and 3 were captured in December 2013 and acquired from City Maps, City of Cape Town. These photographs were captured using Microsoft UltraCam Eagle that recorded the exterior (X, Y, Z) and interior (ω, ϕ, κ) orientation of each digital photograph.

LiDAR was acquired from the Geospatial Services, City of Cape Town and was captured in December 2013 using the ALS50 Airborne Laser Scanner developed by Leica Geosystems based on the Fuguro Fli_Map system. It was recorded in a sinusoidal scan pattern, nominally centred about nadir, achieving an average point density of 1 point/m^2.

4.3.3 Segmentation

The Segmentation Parameters Tuner (SPT) was developed at the Computer Vision Lab (LVC) of the Electrical Engineering Department at the Catholic University of Rio de Janeiro (PUC-RIO). SPT uses a supervised approach to optimises the Baatz segmentation parameters proposed in Baatz and Shäpe (2000) using a Genetic Algorithm (GA) (Davis 1991). The Baatz segmentation

parameter is defined by the weight of the colour, compactness and scale. The GA was used to determine nearly optimal segmentation (parameters and segments) equal to as or similar as possible to, the shape and size of the user defined reference segments (Novack et al. 2011). The level of agreement or fitness between the segmentation and the set of user defined reference segments was measured using a Reference Bounded Segments Booster (Feitosa et al. 2008). The Reference Bounded Segments Booster is formally defined by a mathematical function, and can be found in the SPT User Guide (2008). The methodology to solve SPTs optimization is illustrated in Figure 4.3. The multiresolution segmentation region grow was then applied to merge smaller objects with similar spectral values within individual dwelling extents, for a singular dwelling object.

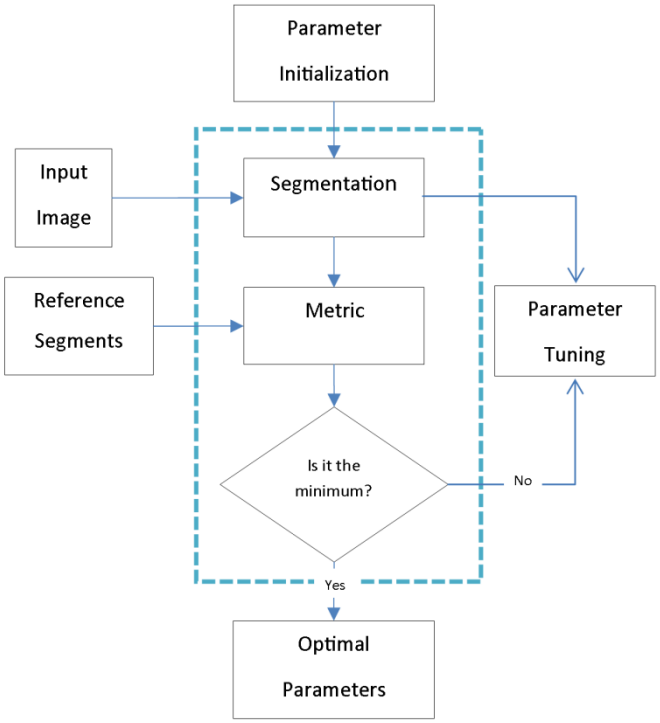


Figure 4.3 SPT optimization methodology.

Source: Achanccaray et al. (2015)

4.3.4 Feature selection and classification

Feature selection aim is to improve classification accuracy by selecting an optimal subset of features (Hapfelmeier & Ulm 2013), as the large number of image object features makes a detailed quantitative exploratory analysis of every individual feature extremely time and labour intensive task (Novack et al. 2008). To compute an optimal subset of features in this study, feature selection was undertaken using the Random Forest algorithm (Kursa, Jankowski & Rudnicki 2010) and the Classification and Regression Trees (CART) (Breiman 1984) on a WV-2 set of features (n=140) and aerial (RGB) set of features (n=60) to evaluate which imagery and features would result in increased classification accuracy. The full lists of features derived per

image are presented in Table 4.2. A glossary defining the features can be found in APPENDIX A.

Table 4.2 Features derived for use in feature selection and classification.

CATERGORY	**TYPE**
Customised	Normalised difference vegetation index (NDVI): (NIR - Red) / (NIR + Red)
Layer values	Mean
	Max difference
Geometry (extent)	Area
	Border length
	Length
	Length/thickness
	Length/width
	Relative border to image border
	Thickness
	Number of pixels
	Volume
	Width
Geometry (shape)	Asymmetry
	Border index
	Compactness
	Density
	Main direction
	Radius of largest enclosed ellipse
	Radius of smallest enclosed ellipse
	Shape index
	Rectangular fit
	Roundness
Texture (GLCM)	Angular second moment
	Contrast
	Correlation
	Dissimilarity
	Entropy

	Homogeneity
	Mean
	Standard deviation
Texture (GLDV)	Angular second moment
	Contrast
	Entropy
	Mean

4.3.4.1 Feature selection with CART

Classification and Regression Trees (CART) proposed by Breiman (1984), is a non-parametric classification approach. A decision tree (De'ath & Fabricius 1999) provides a hierarchical representation of the feature space in which features are allocated to classes based on observations (De'ath & Fabricius 1999). Decision trees are iteratively grown by splitting each node into sub-nodes by finding a best split variable and value until reaching terminal node size (Rokach & Maimon 2005). The Salford Systems software (Salford Systems 2001) was used to implement CART. The Gini index, a measure of heterogeneity was used at each node as a splitting rule, and a 10-fold cross validation was used to determine the optimal tree (Salford Systems 2001). Optimum features were selected based on primary splitters in the decision tree and the ranking of feature importance scores (Laliberte, Browning & Rango 2012). The scores range from 0 to 100 and reflect the contribution of each feature to predict the output class, with a score of 0 reflecting the least favourable contribution, and 100 reflecting optimum contribution.

4.3.4.2 Feature selection and classification using random forest

The Random Forest algorithm (RF) (Breiman 2001) is an ensemble of weak unbiased classification or regression trees (Ismail & Mutanga 2010). RF is based on two techniques namely CART and Bagging (Breiman 1998). Bagging was introduced to reduce the variance while maintaining low bias (Mei et al. 2014). Classification using RF was performed by selecting a bootstrap sample (i.e. with replacement) approximately two-thirds of the original data set. A random subsets possible feature (mtry) is selected with best distribution of data between the nodes of the tree. The final classification is based on simple majority voting determined by all trees in the ensemble (ntree). The remaining one-third of the sample (out-of-bag (OOB)) is used to determine feature importance based on Z-scores and mean prediction error or OOB error. The higher the increase in OOB error indicates that a feature is more important. The results are used for approximations of the classification error and computation of confusion matrices (Kursa et al. 2010). In this study, feature selection and classification using RF was implemented using

Salford Systems software (Salford Systems 2001) which ranked features based on feature importance scores.

4.3.4.3 Rule-based classification with CART

A rule-based classification or membership function classifier uses fuzzy or crisp membership functions and its logical operators to define membership to image objects (Myint et al. 2011). Rule-based classification uses single or several conditions for assigning objects feature to the informal settlement class, and is suited to handle vagueness and ambiguities in information extraction (Rahman & Saha 2008). In order to find relevant image object features and corresponding thresholds, rule-based classification intrinsically relies either on expert consultation (Myint et al. 2011; Kohli et al. 2012), cognitive methods (Zhou et al. 2010; Sebari and He 2013; Belgui et al. 2014) or automatic induction methods (feature selection). The method adopted in this study used the Classification and Regression Trees (CART) to determine the most relevant image object features and the corresponding thresholds for the rule-based classification.

4.3.5 Accuracy assessment

4.3.5.1 Segmentation accuracy

Segmentation algorithms can be evaluated either analytically or empirically (Zhang 1996). Analytical methods assess the algorithm properties and principles, and provide a qualitative assessment of the segmentation (Zhang 1996). Empirical methods measure the property of the segmentation image using an empirical discrepancy method to evaluate the segmentation against an accurate reference (Carleer et al. 2005). Segmentation parameters optimised by the Segmentation Parameter Tuner (SPT) were evaluated using area-based metrics. Area-based accuracy metrics use a set of statistical parameters defined by McGlone and Shufelt (1994) and involves the spatial comparison of an extracted area against a reference area. For this study, the optimised SPT segmentation was compared with reference individual dwellings extents. The true positive (TP) area, false positive (FP) area, false negative (FN) area, completeness and correctness as described in section 3.3.4.1 were used to evaluate segmentation accuracy.

4.3.5.2 Classification accuracy

Classification was undertaken using a disproportionate random stratified sampling scheme (Chen, Wakefield & Lumely 2014). A confusion matrix (Kohavi & Provost 1998) shows the actual and predicted classifications performed by a classification system. Additionally, a KHAT statistic was computed (Congalton & Green 2009) which tests whether the values in the confusion matrix are due to true agreement, or are due to chance agreement.

The McNemar's test (McNemar 1947), which is a non-parametric test based on a 2 x 2 cross tabulation of correctly and incorrectly classified samples, was further used to evaluate any significant difference in accuracy between the RF and rule-based classifiers (Pal & Foody 2010). The McNemar's test is based on a chi-square statistic computed as follows (Leeuw et al. 2006):

$$X^2 = \frac{(f_{12} - f_{21})^2}{(f_{12} + f_{21})^2}$$ Equation 4.1

Where f12 in this study is the number of samples correctly classified by RF, but incorrectly classified by NN, and f21 is the number of samples correctly classified by NN, but incorrectly classified by RF. An x^2 value greater than 3.84% and a statistical significance level of 0.05 indicates a significant difference in performance between RF and NN classification. The McNemar's test has been used in other studies to statistically compare image classification algorithms (Duro et al. 2012; Im et al. 2012).

4.4 Results

4.4.1 Segmentation parameter tuner

Using the Segmentation Parameter Tuner (SPT) on the VHR aerial images, the Genetic algorithm (GA) and search space parameters was defined in the SPT tool. In evolutionary computing terms (SPT User Guide 2008), the number of individuals within a population represents the potential solutions for the optimised segmentation parameters. By increasing the population size (n=20) and number of generations (n=50) (evolution cycle), a greater genetic variety is introduced that assist to determine optimal segmentation solutions. To further increase the algorithm's inherent random quality, the number of experiments (n=10) was selected. Overall the goal of the GA is to converge to a (nearly) optimum solution. It was found by increasing the procedure parameters such as the number of experiments, it resulted in overestimation (SPT User Guide 2008), with significant changes in the output segmentation parameters. Segmentation search parameters were optimised between a minimum and maximum range, for the Baatz segmentation procedure proposed in Baatz and Shäpe (2000). Using SPT, the optimally tuned segmentation parameters were determined for test site 1, 2 and 3. The SPT tool was used to find optimal segmentations parameters for each informal settlement, as inter-informal settlements dwelling characteristics was too diverse to apply the segmentation parameters set from test site 1, the training area to test site 2 and 3, the test areas. The highest scale parameter was achieved by test site 2, followed by test site 3 and test site 1. The change in scale demonstrates the size difference in dwellings across the three informal settlements, with test site 2 dwellings having larger dwelling dimensions and test site 1 the smallest.

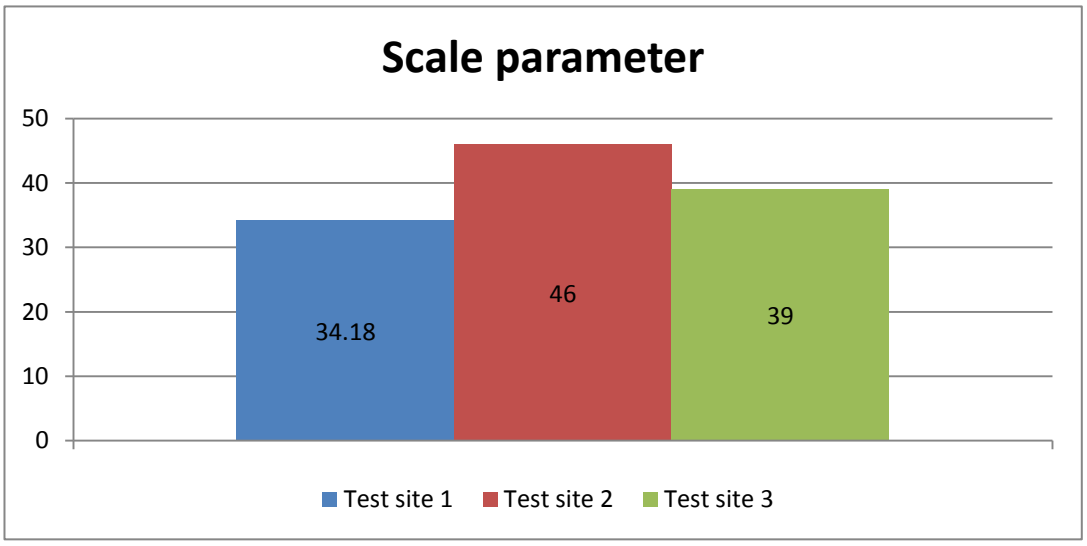


Figure 4.4 The optimised scale parameters derived from SPT for test sites 1, 2, and 3.

According to Li and Liu (2015), homogeneity is controlled by the *weight of colour* and *compactness*, and the sum of the coefficients for each pair should equal to 1. From this study it was found not always to be the case in the SPT tool. The *weight of colour* was similar in test site 1, 2 and 3 (~0.12) showing the spectral information was relatively not important (relative to 1), in comparison to the *weight of compactness* (~0.8), which determines the spatial homogeneity of the segmentation (eCognition 2012).

Table 4.3 Results obtained with the segmentation parameter tuner for the different informal settlements.

	Test site 1	Test site 2	Test site 3
Weight Colour	0.124	0.123	0.12
Weight Compactness	0.806	0.814	0.824
Fitness	0.235	0.226	0.073

After the optimal segmentation parameters were derived using the aerials imagery, the segmentation was overlaid on the 8-band WV-2 imagery to extract the spectral information for classification. In Figure 4.5 it can be seen that the high resolution aerial imagery closely represents the informal dwelling extents, however when the segmentation is overlaid on the WV-2 imagery, the informal dwelling extents are not as clear due to the change in spatial resolution.

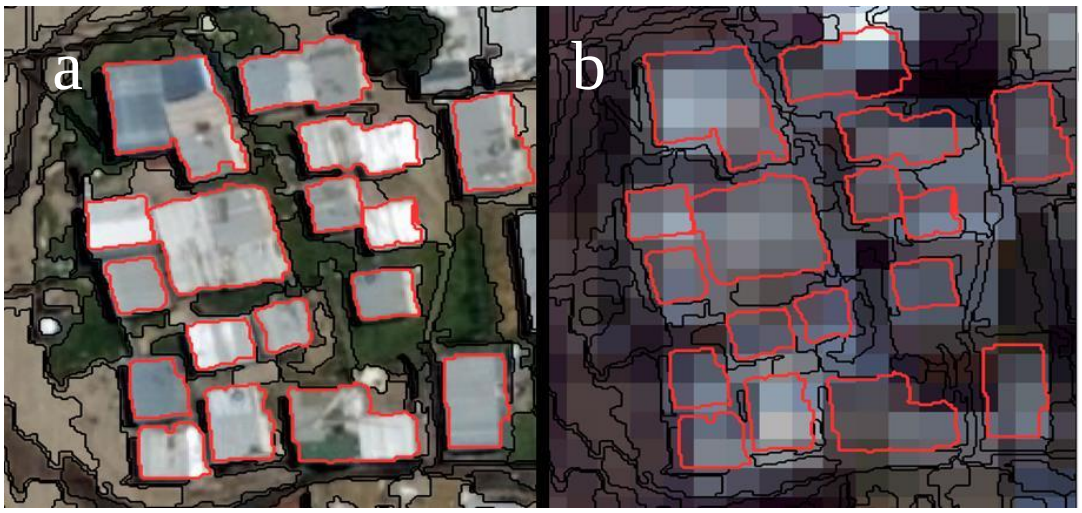


Figure 4.5 The segmentation parameters derived using SPT tool on the VHR aerial imagery (a), and applied to the 8-band WorldView-2 imagery (b).

4.4.2 Area-based segmentation accuracy

The results of the area-based accuracy assessment for the derived SPT segmentation parameters are shown in Figure 4.6. A reference area for test site 1 (67.7 m^2), test site 2 (64.32 m^2) and test site 3 (155.49 m^2) indicate the difference in the area and size of the dwellings in the respective informal settlements. The highest extracted building area was achieved by test site 2 (170.70 m^2), followed by test site 1 (63.45 m^2) and test site 3 (61.02 m^2). Test site 2 achieved the highest TP (147.90 m^2) followed by test site 1 (59.83 m^2) and test site 3 (57.93 m^2). The lowest FP or omitted area was recorded for test site 3 (7.12 m^2), followed by test site 2 (7.59 m^2) and test site 1 (7.88 m^2). However, the highest FN or included area was achieved by test site 2 (22.80 m^2), with test site 3 and test site 1 achieving an area of 3.11 m^2 and 3.62 m^2 respectively. To ensure high accuracy is achieved, the TP area should be as close to the reference and extracted areas as possible (such as test sites 1 and 3) with low FP and FN area, indicating low omission and low inclusion. However when lower accuracy is achieved, the extracted area is generally greater than the reference area (such as test site 2), with high FP and FN areas indicating high omission and inclusion. In test site 1 and test site 3, the TP, FP area and FN area are similar, showing the consistency of the SPT segmentation parameters in the two informal settlements when dwelling dimensions are similar.

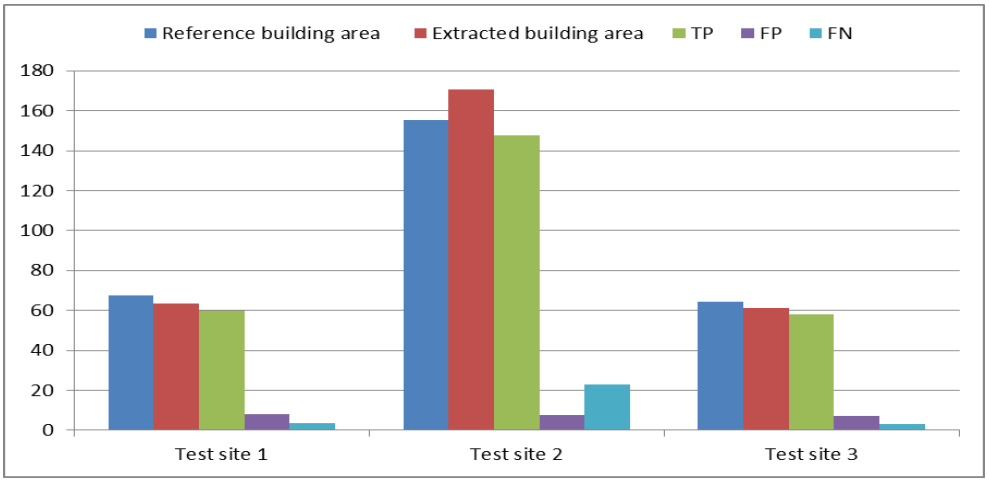


Figure 4.6 Area-based accuracy assessment for test site 1, 2 and 3 based on 50 dwellings.

To further understand the relationship of TP, FN and FP, completeness and correctness was calculated. In this study, the highest completeness was achieved by test site 3 (95%), followed by test site1 (94%) and test site 2 (86%). However the highest correctness was achieved by test site 2 (95%) followed by test site 3 (89%), and test site 1 (88%). From the results as seen in Figure 4.7, it was noted that that high FN area degraded completeness, whereas high FP area degraded correctness (Hermosilla 2011).

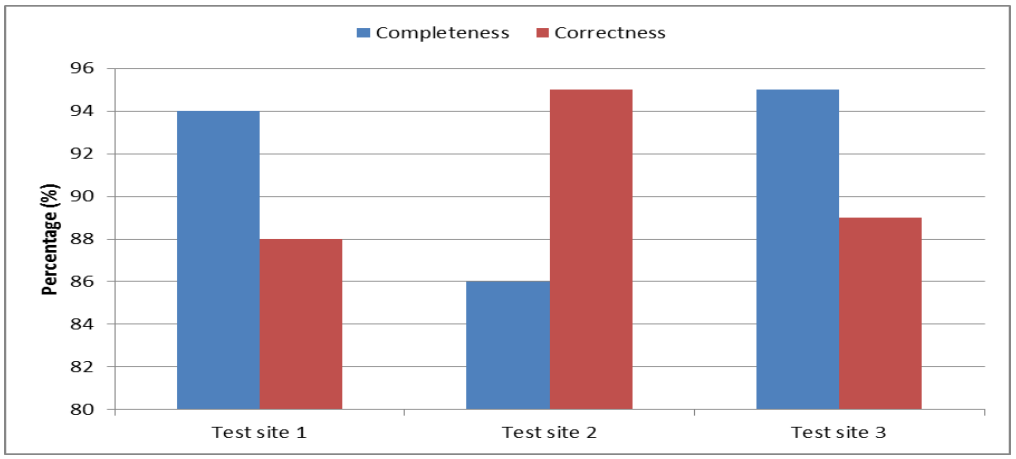


Figure 4.7 Completeness and correctness computed for test site 1, 2 and 3.

The relatively high completeness and correctness percentages was achieved in all three informal settlements, showing that the SPT tool is efficient at deriving segmentation parameters that do not include ‘false’ areas (FN and FP), as it well understood that the accuracy and reliability of classification largely depends on the accuracy of the segmentation (Baatz and Schäpe 2000; Benz et al. 2004).

4.4.3 Feature selection using RF on WV-2 imagery

Using the RF algorithm the original object feature set (n = 140) was reduced to a subset of 108 features which comprised geometry features (n = 21), texture features (n = 78) and mean layer value features (n = 9). Table 4.4 shows just the 30 highest ranked features in the feature subset.

Table 4.4 Top 30 ranked features, based on the RF feature selection algorithm, on the WV-2 imagery.

Ranking	Feature	Importance Score (out of 100)
1.	Roundness	100
2.	Elliptical fit	93.31
3.	Density	92.61
4.	Shape Index	92.30
5.	Border Index	91.49
6.	GLCM Mean Green	88.04
7.	Compactness	87.69
8.	Mean nDSM	75.31
9.	Rectangular Fit	74.87
10.	Mean Blue	65.94
11.	Radius of largest enclosed ellipse	56.59
12.	Mean Yellow	51.56
13.	Asymmetry	49.53
14.	Mean Red	48.18
15.	Area	48.12
16.	Mean Red Edge	45.14
17.	Mean Pan	45.14
18.	GLCM Mean Near infrared	43.77
19.	Mean Green	43.32
20.	GLCM Mean Red	39.57
21.	Volume	36.34
22.	Number of pixels	36.19
23.	Length	34.66
24.	GLCM Mean Blue	32.79
25.	Width	32.63
26.	GLCM Contrast Red	30.64
27	Border length	29.62

28.	Mean Near infrared	28.11
29.	GLCM Entropy Red	25.59
30.	Length/Width	23.76

It is interesting to note the selection of the features in the top 10, 20 and 30 when compared to the features in the reduced subset and full set of features, as seen in Figure 4.8. In the reduced and full set of features, texture was dominant but slowly was found to be less important with the increase in rank, with 78 texture features in the subset but only 1 and 2 features in the top 10 and 20 respectively. In contrast, geometry features became more important with the increase in rank, with 7, 10 and 15 features of the top 10, 20 and 30 and the top five most important features. In addition the importance and frequency of layer value features fluctuated with 8 features included in the top 20 but only 2 in the top10, with the highest important positions of 8th and 10th.

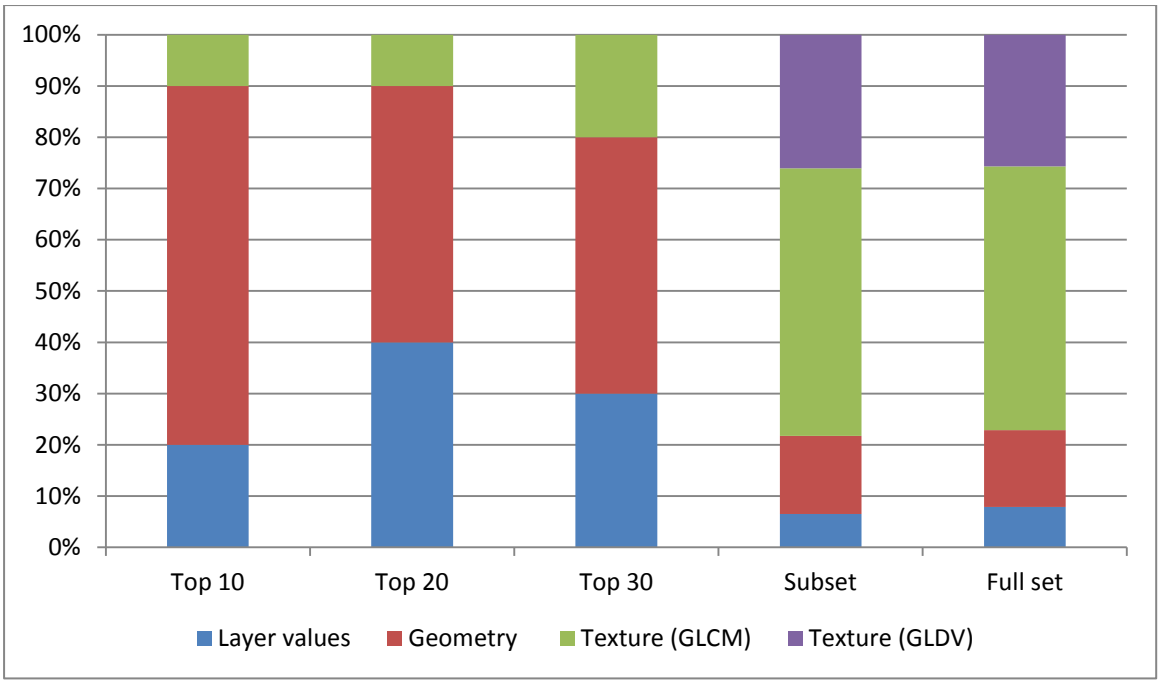


Figure 4.8 Frequency of occurrence of features from the WV-2 imagery in the top rankings of the RF feature selection.

4.4.4 Feature selection using RF on aerial imagery

Using the RF algorithm the original object feature set (n = 60) was reduced to a subset of 28 features which comprised geometry features (n = 20), texture features (n = 4) and mean layer value features (n = 4). Table 4.5 shows the 30 highest ranked features.

Table 4.5 Top 30 ranked features, based on the RF feature selection algorithm, on the aerial imagery.

Ranking	Feature	Importance Score (out of 100)
1.	Mean Blue	100.00
2.	GLCM Dissimilarity Coastal Blue	87.67
3.	Mean Red	84.83
4.	Elliptical Fit	80.21
5.	GLCM Dissimilarity Panchromatic	78.99
6.	Mean nDSM	74.74
7.	Mean Green	63.51
8.	GLCM Entropy Panchromatic	60.60
9.	Rectangular Fit	58.31
10.	Density	57.32
11.	Radius of largest enclosed ellipse	53.70
12.	Shape Index	51.20
13.	GLCM Standard Deviation Panchromatic	48.89
14.	Roundness	47.53
15.	Radius of smallest enclosed ellipse	42.47
16.	Compactness	37.99
17.	GLDV Contrast Panchromatic	35.22
18.	Border Index	27.82
19.	Asymmetry	14.80
20.	Number of pixels	12.50
21.	Volume	12.08
22.	Area	7.55
23.	Border Length	6.92
24.	Width	4.98
25.	Length	4.34
26.	Relative border to enclosed border	3.91
27	Length/Width	3.13
28.	Main Direction	1.06
29.	Thickness	0.00
30.	Length/Thickness	0.00

A very different set of features were selected from the aerial imagery when compared to the features selected from the WV-2 imagery. It can be noted that there is less texture features in the aerial imagery due to only 3 bands when compared to WV-2 9 bands. Nonetheless, texture features increased in importance from 5 in the top 20 to 3 in the top 10, as seen in Figure 4.9 with positions 2, 5 and 8 in the top 10. In contrast geometry features frequency decreased with the increase in importance with 21, 11 and 3 in the top 30, 20 and 10. On the other hand, mean layer features frequency stayed constant with 4 features in top 10, 20, and 30, with all features in the top 7 with rankings 1, 3, 6, and 7.

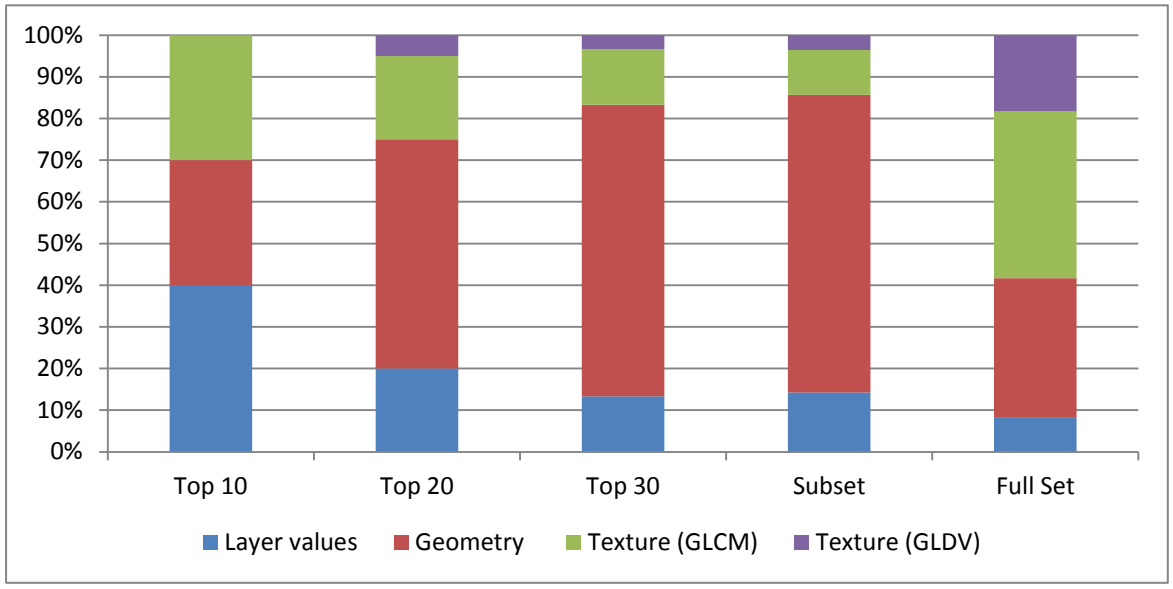


Figure 4.9 Frequency of occurrence of features from the aerial imagery in the top rankings of the RF feature selection.

The WV-2 and aerial imagery features were combined in one feature set (n=200) and the RF feature selection was undertaken to assess what amount of WV-2 and aerial features would be in the top 10, 20, 30. A subset of 136 features was selected with 42 aerial and 94 WV-2 features. A large proportion of features were derived from the aerial imagery with 9, 16 and 20 features in the top 10, 20 and 30, and more specifically the geometry features with 7, 12 and 16 features in the top 10, 20 and 30 as seen in Figure 4.10. The 20 highest ranked features from the WV-2 and aerial imagery combined feature set are shown in Table 4.6.

Table 4.6 Top 20 ranked features, based on the RF feature selection algorithm, on the WV-2 and aerial imagery.

Ranking	Feature	Importance Score (out of 100)
1.	Compactness	100.00
2.	Mean Panchromatic (WV-2 imagery)	99.66

3.	Shape Index	94.97
4.	Elliptical Fit	93.29
5.	Density	91.58
6.	Roundness	89.41
7.	Border Index	86.49
8.	Mean nDSM	85.94
9.	Mean Green (aerial imagery)	71.33
10.	Number of pixels	70.33
11.	Mean Blue (aerial imagery)	67.17
12.	Area pixel	65.42
13.	Rectangular fit	63.55
14.	Mean Red (aerial imagery)	63.04
15.	GLCM Mean Blue (WV-2 imagery)	62.63
16.	Mean Red (WV-2 imagery)	61.76
17.	Border Length	58.72
18.	Number of pixels	58.12
19.	Volume	56.56
20.	Mean Yellow (WV-2 imagery)	53.41

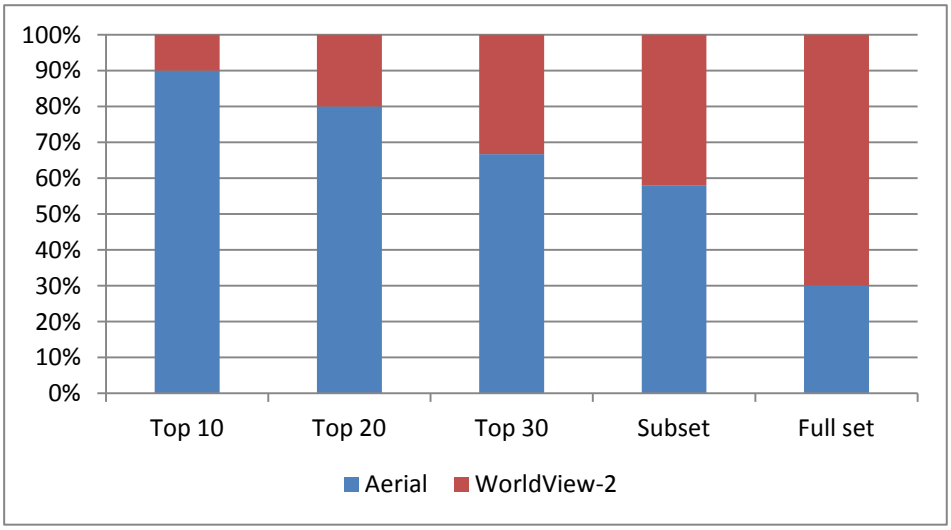


Figure 4.10 Frequency of occurrences of features from the WV-2 and aerial imagery in top rankings of the RF feature selection.

4.4.5 RF classification

Classification using the full set of WV-2 features (n=140) and the subset (n =108) in the test site 1 resulted in an overall accuracy of 96.67%, and when tested in test site 2 and test site 3 it achieved 75% and 94.72% as shown in Figure 4.11. However a decrease in accuracy can be

noted for test site 1 in top 10 (92.86%), top 20 (90.53%) and the top 30 (92.86%) when compared to the full and subset of features. Likewise test site 3 a similar trend and results were achieved with the full and subset of features achieving 98.33% and 96.68%, and a decrease in accuracy in the top 30 (95%), top 20 and 10 (94.72%). In contrast, test site 2 achieved an increase in overall accuracy with fewer features, with the full and subset of features achieving 75% compared to top 30(78.33%) , top 20 (78.33%) and top 10 (90%) respectively.

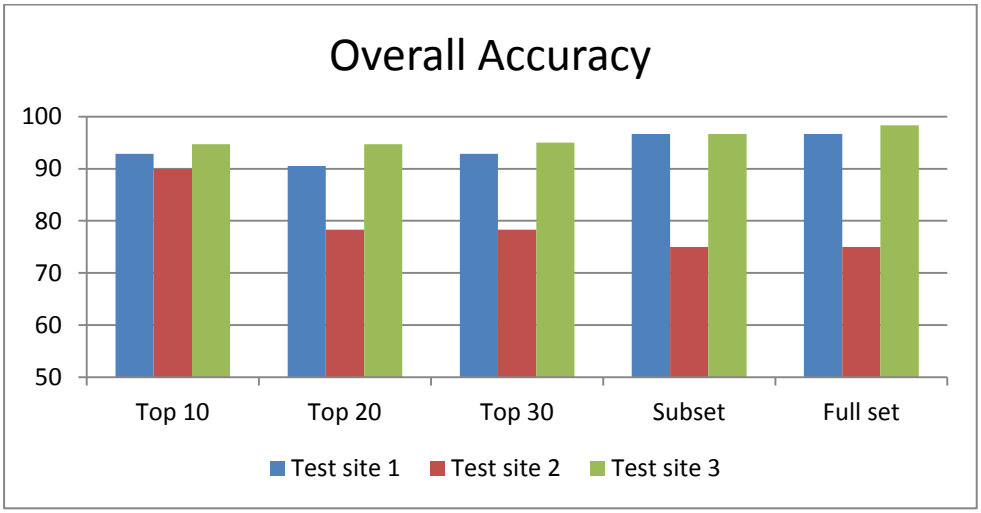


Figure 4.11 Overall accuracy percentages achieved for WV-2 imagery in test site 1, 2 and 3.

Classification using the full set of aerial features (n=60) and the subset (n =28) in the test site 1 resulted in an overall accuracy of 98.33%, and when tested in test site 2 and test site 3 it achieved 78.33% and 96.41% as shown in Figure 4.12. A general trend can be observed in test site 1 and test site 3 with the selection of fewer features resulting in lower accuracy in contrast to test site 2.

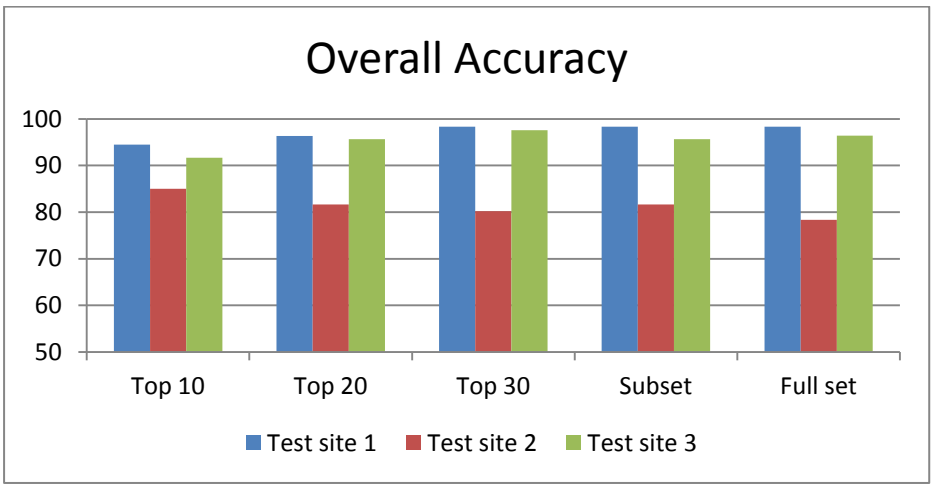


Figure 4.12 Overall accuracy percentages achieved for aerial imagery in test site 1, 2 and 3.

When the WV-2 and aerial imagery features were combined in one feature set (n=200) the highest overall accuracy was achieved in the top 30 for test site 1(97.88%) and test site 3 (96.86%), whereas the highest overall accuracy was achieved was achieved in the top 10 by test

site 2 (90%) as seen in Figure 4.13. Similarly a trend can be observed with the selection of more features resulting in higher accuracy in test site 1 and 3.

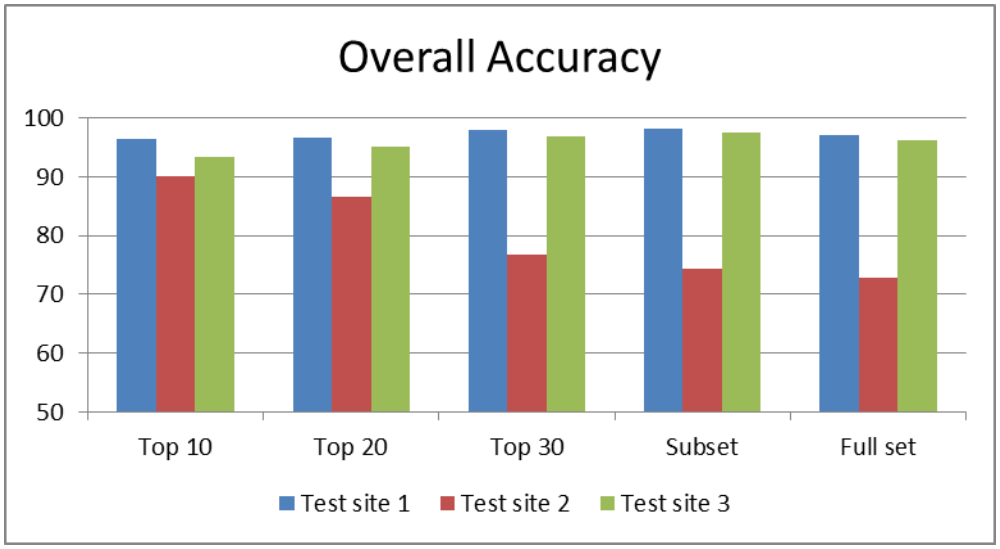


Figure 4.13 Overall accuracy percentages achieved for WV-2 and aerial imagery in test site 1, 2 and 3.

Overall when comparing all the 3 feature sets (WV-2, aerial, WV-2 and aerial), in test site 1 the combined WV-2 and aerial features achieved the highest accuracy in the top 10 (96.33%), followed by aerial (94.51%) and WV-2 (92.86) features. Similar a trend can be observed in top 20; however in the top 30, subset and the full set, aerial imagery features achieved the highest overall accuracies followed by the combined features set and lastly the WV-2 features.

In test site 2, the combined feature set and the WV-2 features achieved the same overall accuracy (90%) followed by the aerial features (85%) in the top 10. However, the combined set of features achieved the highest accuracies in the top 20 (86.67%), followed by the aerial features (81.67%) and WV-2 features (78.33%). In the top 30, the subset and full feature set, the aerial features achieved the highest overall accuracies (80.23%, 81.67%, 78.33), followed by the combined feature set (76.67%, 74.37%, 72.77%) and WV-2 features (78.33%, 75%, 75%).

The WV-2 features achieved the highest overall accuracy in the full set of features (96.41%) in test site 3 followed by aerial features (96.41%) and the combined feature set (96.16%). In addition the WV-2 features achieved the highest accuracy in the top 10. The top 20 and top 30 highest accuracies were achieved by the aerial features, followed by the combined feature set and WV-2 features. Lastly the combined feature set achieved the highest accuracy in the subset of features (97.45%), followed by the WV-2 features (96.68%) and the aerial features (95.68%).

4.4.6 Feature selection using CART on WV-2 and aerial imagery

Using the CART algorithm the original WV-2 feature set (n=140) was reduced to a subset of 6 features which comprised geometry features (n =5), texture features (n=0) and mean layer value features (n =1). The feature importance scores are shown in Table 4.7. It can be noted that none of the 9 WV-2 bands and any derived texture features were included in the feature subset in contrast to the features selected by the RF algorithm. However similarly to the features selected by the RF algorithm, the CART feature subset is mostly geometry features.

Table 4.7 The highest ranked features based on the CART algorithm, on the WV-2 imagery.

Ranking	Feature	Importance Score (out of 100)
1.	Mean nDSM	100.00
2.	Density	93.54
3.	Elliptical Fit	87.11
4.	Rectangular Fit	81.09
5.	Radius of largest enclosed ellipse	81.09
6.	Compactness	74.74

The CART algorithm was also used to reduce the original aerial feature set (n=60) to a subset of 5 features which comprised geometry features (n = 1), texture features (n = 0) and mean layer value features (n = 4). As shown in Table 4.8, the three aerial imagery bands and the nDSM was selected and ranked with the same feature importance. In addition no texture features were selected in the final subset.

Table 4.8 The highest ranked features based on the CART algorithm, on the aerial imagery.

Ranking	Feature	Importance Score (out of 100)
1.	Mean nDSM	100.00
2.	Mean Green	100.00
3.	Mean Blue	100.00
4.	Mean Red	100.00
5.	Density	93.54

4.4.7 Rule-based classification using WV-2 and aerial imagery feature set

Rule-based classification was undertaken using the reduced WV-2 feature subset and aerial feature subset as illustrated in Figure 4.14. The highest overall accuracy using the WV-2 feature subset was achieved in test site 1 with 95.67% followed by test site 3 (93.11%) and test site 2 (75.77%). Similarly using the aerial feature subset test site 1 achieved the highest overall accuracy (96.85%), followed by test site 3 (94.45%) and test site 2 (74.33). Overall the highest accuracy achieved using CARTs reduced feature subsets in test site 1 and test site 3 was the aerial features (96.85% and 94.45%), with the WV-2 features achieving the highest accuracy in test site 2 (75.77%).

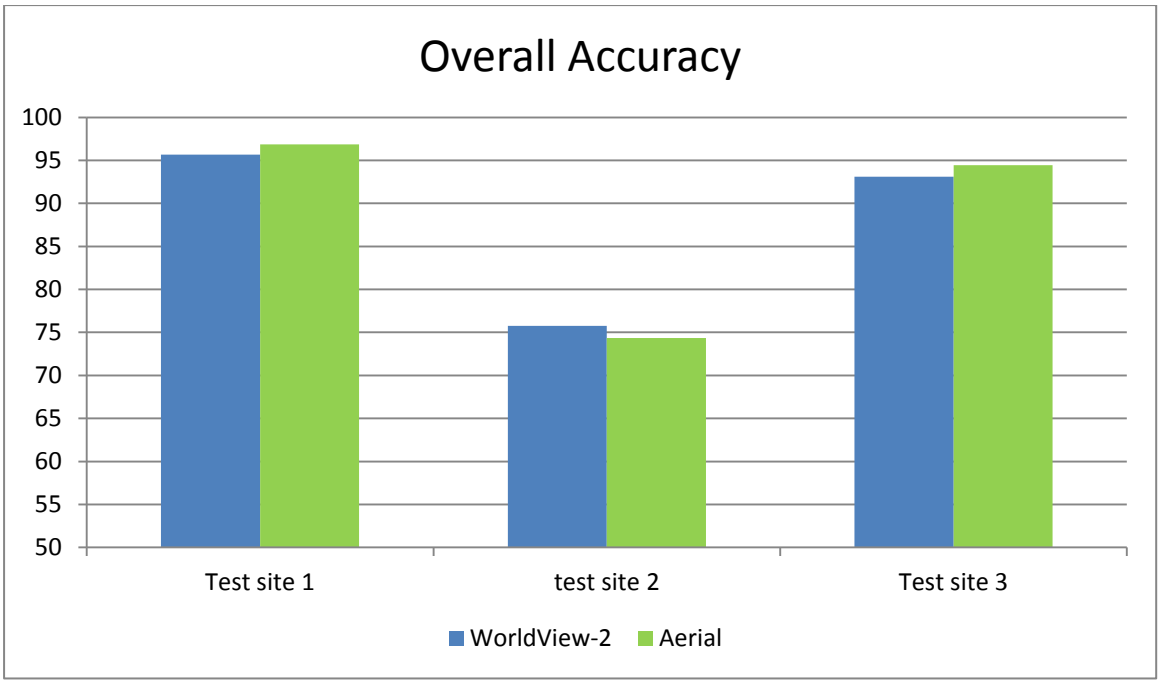


Figure 4.14 Overall accuracy percentages using CARTs WV-2 and aerial reduced feature subsets in test site 1, 2 and 3.

4.5 Discussion

In this study a dynamic methodology to identify individual informal dwellings using WorldView-2 imagery, aerial imagery and LiDAR datasets within an OBIA environment was developed. This study represents a first attempt at mapping individual informal dwellings using an objective identification approach for segmentation, feature selection and classification. The results of this study are significant within a South African and global context, given the urgent need for reliable spatial data for the evaluation of performance and intervention based programs (Kit 2013; Kohli et al. 2013). The following sections discuss the results in more detail.

4.5.1 Segmentation

It was important to determine optimal segmentation parameters, as it influences the reliability of the classification results (Navulur 2007; Zhang et al. 2010; Gao et al. 2011; Sebari and He 2013). The SPT tool was used to optimise the Baatz segmentation parameters including the *scale*, the *weight of colour* and *compactness*, instead of a subjective time-consuming and trial-by-error process (Belgui et al. 2011). The output of SPT was optimal segmentation parameters and image objects that represented the lowest discrepancy, or the highest agreement between reference and producers segments (Novack et al. 2012). The SPT parameters and image objects were optimised on the aerial imagery (12.5 cm) and subsequently overlaid on the WorldView-2 imagery (0, 46 m and 1.84 m). This specific method was implemented to take advantage of the spectral resolution of the WorldView-2 imagery, and the spatial resolution available in the aerial imagery, as no current space borne sensor (commercially available) can achieve a comparable spatial resolution. As a result, this allowed individual informal dwellings within the test site 1, 2 and 3 informal settlements to be optimally segmented, as indicated by the high completeness and correctness values. The highest completeness and correctness values were achieved by test site 3 (95% and 98%), followed by test site 1 (94% and 88%) and test site2 (86% and 95%). From the results, it was noted that that high FN area degraded completeness, whereas high FP area degraded correctness (Hermosilla 2011).

The increased completeness and correctness achieved in test site 3 is linked to its informal dwelling characteristics. The dwellings in test site 3 are characteristically similar in terms of building dimensions when compared to the dwellings in test site 1and test site 2, and is either linked to the initial method of occupation of the land (‘land grab method’), or the availability of land. Correspondence in the respective informal settlements with the local management and street committees confirmed that fires and flooding in neighbouring informal settlements forced many to relocate at once to test site 3, and divide land accordingly. This resulted in dwellings having similar dimensions in test site 3 in contrast to test site 1 and test site 2 that gradually became occupied over time, with older dwellings having greater dimensions compared to the newer dwellings. Test site 3 thus achieved superior completeness and correctness, as dwellings dimensions were more consistent and less variable in relation to SPT user defined segments than in test site 1 and test site 3. The results of this study address the recommendations by Novack et al. (2011) to perform an expanded analysis to assess the influence of WorldView-2 on the segmentation process. It was found that the WorldView-2 spatial resolution was too coarse to determine optimal segmentation parameters for informal dwelling mapping, and that additional VHR aerial imagery was required.

4.5.2 Feature selection using RF

To undertake a detailed quantitative exploratory analysis of every object feature (spectral, spatial and textural) is a time consuming and labour-intensive task (Novack et al. 2011; Laliberte, Browning & Rango 2012).The addition of redundant and unnecessary features can lead to the deterioration of classification accuracy known as Hughes phenomena, and can be avoided by reducing the data dimensionality of the given feature set (Hughes 1968). There are two approaches to reduce data dimensionality, either 1) the minimum set of features can be defined or, 2) all relevant features can be defined (Nilsson et al. 2007). In this study, CART found the minimum set of features whereas RF found all relevant features.

Using the Boruta algorithm the original WV-2 feature set (n=140) and original aerial feature set (n=60) was reduced by 23% and 53% respectively. In the feature subsets, WV-2 imagery mainly selected texture features (n=78) followed by geometry features (n=21), whereas in contrast aerial imagery mainly selected geometry features (n=20) followed by mean layer features (n=4).The inclusion of textural features in both feature subsets was not surprising given the texturally diverse informal settlement environment, and the significance of texture in previous studies (e.g. Kohli et al. 2013). In addition the inclusion of all WV-2 bands is consistent with the findings of Novack et al. (2011) who found that the additional WV-2 bands (Coastal Blue, Yellow, Red-Edge, NearIR-2) comprised more than 50% of the WV-2 bands in the final feature subset.

A shift in the features selection by RF can be noted when comparing the respective feature subsets of the WV-2 features to the features in the top 30. A significant decrease in the WV-2 textures features can be observed from 78 features in the feature subset (72% contribution) to 6 features in the top 30 (20% contribution). Likewise an increase in geometry features can be seen from 21 features in the feature subset (19% contribution) to 15 features in the top 30. There were no significant changes in the selection of features between the aerial feature subset (n=28) and the top 30 features, with the inclusion of one additional geometry and texture feature.

On closer examination of the respective features selected, a trend can be observed from the feature subset to the features in the top 30, 20 and 10. The WV-2 imagery selected less texture features from 78% contribution in the feature subset (n=108) to 10% in the top 10, and from 21 geometry features in the feature subset (19% contribution) to 7 geometry features in the top 10. Likewise a similar trend can be noted in feature selection of the aerial imagery, with a noteworthy increase in texture features in the top 10 (30% contribution) when compared to the

11% contribution in the feature subset (n=28). Similarly the selection of geometry features decreased with only 3 features in the top 10 and 20 features in the feature subset.

Interestingly when examining the combined WV-2 and aerial features (n=200), a subset of 136 features was selected with 42 aerial and 94 WV-2 features. It was found that a greater distribution of aerial features is present in the top 30, 20 and 10. In the top 30, a total of 20 aerial features were selected , in comparison to the 9 aerial features selected in the top 10. In the top 10, the highest ranking WV-2 feature was the mean panchromatic band (WV-2), with an importance score of 99.66 closely behind compactness (100), which was derived from the aerial imagery.

4.5.3 Feature selection using CART

Using the CART algorithm the original WV-2 feature set (n=140) and original aerial feature set (n=60) was reduced by 95% and 91% respectively. In both the WV-2 and aerial feature subsets no texture features was selected, and is interesting given its inclusion by the RF algorithm. Furthermore, none of the WV-2 bands were included in the WV-2 feature subset, in contrast to all the aerial bands that were selected in the aerial feature subset. The nDSM was selected and ranked with high importance (100) in both the WV-2 and aerial imagery feature subsets, indicating its significance as an auxiliary dataset over other mean layer and geometry features. In addition, the use of an nDSM for informal settlement mapping has not been explored in previous literature, and is promising in the context of this study. Thus this study is the first to include nDSM to assist classification of informal dwellings within an informal settlement, and fulfils a recommendation by Tiedi et al. (2010) and Pinho et al. (2011) towards defining a model that can differentiate between bare soil and metal roof structures. Additionally the inclusion of an nDSM can assist to remove any surface vegetation that has added to classification problems as noted in Kohli et al. (2012). Different feature subsets have been selected by the RF and CART algorithm for the WV-2 and aerial imagery, and it will be interesting to see how the different feature selection subsets classification will perform when compared to each other, as the RF algorithm selected a subset of 108 WV-2 features and 28 aerial features, whereas CART selected a subset of 6 WV-2 features and 5 aerial features.

4.5.4 RF classification

Random forest classification was undertaken on various WV-2 and aerial feature subsets (i.e. top 10, 20, 30, subset and full set) to assess what is the impact of different feature subsets on the classification accuracies in test site 1, 2 and 3. Using the full WV-2 feature set (n =140) and the full aerial feature set (n =60), the highest overall accuracy was achieved by the aerial features in test site 1 (98.33%) and test site 2 (78.33%), however WV-2 achieved the highest accuracy in

test site 3 (98.33%). When using the reduced aerial feature subset (n=28) no significant difference was noted in test site 1, however an increase in accuracy can be seen in test site 2 and a decrease in accuracy in test site 3, with the increase in geometry features in the feature subset. Likewise there was no change in classification accuracy in test site 1 with the WV-2 feature subset (n=108) but a decrease in classification accuracy in test site 3 (96.68%).

In the top 30 features, both the WV-2 and aerial subsets mostly consisted of geometry features, with less WV-2 texture features (n=6) and more aerial features (n=5) selected when compared to the feature subsets (n= 78 and n = 4). The decline of WV-2 texture features in the top 30 resulted in a decrease in classification accuracy in test site 1 (92.86%) and test site 3(95%) but an increase in accuracy in test site 2. In contrast the aerial texture features resulted in an increase in classification accuracy in test site 3 (97.68%), and a decrease in test site 2 (80.23%).

In the top 10 and top 20, contrasting trends can be observed in the selection of WV-2 and aerial features. In the WV-2 feature sets, less texture and mean layer features are selected from top 20 (n=12 and n=8) to top 10 (n=1 and n=2), with the increase in geometry features from 10 features in the top 20 to 7 features in the top 10. Similarly in the aerial feature sets, more texture and mean layer features were selected in the top 10, with decrease in geometry features from 11 features in the top 20 to 3 features in the top 10. The apparent difference of the WV-2 and aerial features subsets however resulted in similar overall classification accuracies, with the aerial features in the top 10 achieving the highest accuracy in test site 1 (94.51%) and the WV-2 features in test site 2 (90%) and test site 3 (94.72%) as seen in Table 4.9.

Table 4.9 The overall classification accuracies for the RF algorithm in test site 1, 2 and 3.

		Top 10 (%)	**Top 20 (%)**	**Top 30 (%)**	**Subset (%)**	**Full Set (%)**
WV-2	Test site 1	92.86	90.53	92.86	96.67	96.67
	Test site 2	90.00	78.33	78.33	75.00	75.00
	Test site 3	94.72	94.72	95.00	96.68	98.33
Aerial	Test site 1	94.51	96.34	98.33	98.33	98.33
	Test site 2	85.00	81.67	80.23	81.67	78.33
	Test site 3	91.67	95.68	97.54	95.68	96.41
Combined WV-2 & aerial	Test site 1	96.33	96.54	97.88	98.11	97.00
	Test site 2	90.00	86.67	76.67	74.37	72.77
	Test site 3	93.33	95.11	96.86	97.45	96.16

Another trend can be seen in the overall results of test site 1, 2 and 3 when inspecting the WV-2 and aerial feature sets. Regardless of the imagery used, a steady decrease in classification accuracy in test site 1 and test site 3 can be noted from the full feature subset to the top 10 features and the inverse for test site 2. For example, the WV-2 classification accuracy decreased from the full feature subset (n=140) in test site 1(96.67%) and test 3 (98.33%) to 92.86% and 94.72% in the top 10 respectively. Likewise a similar decrease in classification accuracy, irrespective of the features selected can be observed for the full aerial feature set (n=60) in test site 1(98.33%) and test site 3(94.41%) to 94.51% and 94.67% in the top 10. Moreover, an increase in classification accuracy can be noted in test site 2 from the full feature subsets of WV-2 (75%) and aerial imagery (78.33%) to the top 10 features (90% and 85%). Thus in test site 1 and 3, the feature subsets achieved the highest classification accuracy, and with the decrease in the number of features, it resulted in a decrease in accuracy. However in test site 2, the increase in the number of features, resulted in an increase in classification accuracy

The reason behind the contrasting trends, regardless of the WV-2 and aerial features selected, is due similar dwelling characteristics (spatially and spectrally) in test site 1 and 3 when compared to test site 2 as outlined in Section 1.5. Test site 2 is notably different from test site 1 and 3, as the dwelling roofs are covered in black sheet plastic and not corrugated iron, and the dimensions of the dwellings are significantly larger than test site 1and 3 and not similar within test site 2. From the classification accuracies it can be observed, that a smaller feature set resulted in higher accuracies and thus requires fewer features to characterise a seemingly complex informal settlement.

A similar trend can be noted in the combined WV-2 and aerial feature set (n=200), with the decrease in overall classification accuracies in test site 1 and 3 and increase in test site 2, with the decrease in the number of features selected. In the feature subset (n= 136), 94 features were WV-2 features (69% contribution) and 42 were aerial features (31 % contribution) which resulted in high overall accuracies in test site 1(98.11%) and test site 3(97.45%). In the top 10, 20 and 30 features, less WV-2 features and more aerial features were selected, with 9 aerial features in the top 10. This resulted in decrease in classification accuracy in test site 1 (96.33%) and test site 3(93.33%) but a significant increase in accuracy in test site 2 from 72.77% to 90%. Similarly to the WV-2 feature subset and aerial feature subset, the combined WV-2 and aerial feature subset achieved the highest overall accuracy in test site 1 and 3, with the top 10 features in all three feature sets (WV-2, aerial and combined WV-2 and aerial features) achieving the highest accuracy in test site 2.

4.5.5 Rule-based classification using CART feature set

A rule based classification was undertaken using the reduced WV-2 feature subset (n=6) and aerial feature subset (n=5). The WV-2 feature subset mostly consisted of geometry features (n=5) with no texture or WV-2 bands, which resulted in high overall accuracies in test site1 and 3, and a decrease in classification accuracy in test site 2. Likewise the aerial feature subset which consisted of mainly mean layer features (n=4) achieved comparable classification accuracies in test site 1, 2 and 3 as shown in in Table 4.10. It can be noted that the results in test site 1, 2 and 3 are very similar, and is related to the high importance scores given to mean nDSM in both feature subsets (n=100). In addition, the similarity of the overall accuracies achieved can be related to the similarities in dwelling characteristics in test site 1 and 3 in comparison to test site 2 as highlighted in previous sections.

Table 4.10 The overall accuracies for the CART algorithm in test site 1, 2 and 3.

	Test site 1 (%)	Test site 2 (%)	Test site 3 (%)
WV-2	95.67	75.77	93.11
Aerial	96.85	74.33	94.45

Contrary to the classification accuracies achieved by the RF algorithm in the top 10 in test site 2, a small feature subset selected by CART did not achieve higher accuracies than test site 1 and 3 as anticipated from the RF results. Classification accuracies from the subsets selected by the CART and RF algorithm revealed that there is no specific or best suited feature(s) for mapping informal dwellings as classification accuracies were similar using different feature subsets, as the WV-2 (n=138), aerial (n=28), and the combined WV-2 and aerial features (n=108) in the feature subsets achieved similar accuracies in test site 1 (96.67%, 98.33% and 98.11%) , in test site 2 (75% 81.67% and 74.37%), and test site 3 (96.68%, 95.68% and 97.45%) even though different features were selected (i.e. geometry, texture and mean layer features). In addition similar classification accuracies were achieved in the top 30, 20 and 10 regardless of the WV-2, aerial, and combined WV-2 and aerial features selected. Likewise when examining the results achieved by CART, comparable classification accuracies were achieved in test site 1, 2 and 3 but it however can be attributed to the high importance score given to the mean nDSM in both the WV-2 and aerial feature subsets unlike the different feature importance scores given in the RF subsets.

The varying features selected by the RF and CART algorithm and indifferent classification accuracies achieved, demonstrate the challenge of mapping informal settlements and in testing the methods transferability to other areas (Salehi et al. 2011; Pinho et al. 2012). Hofmann,

Blaschke, & Strobl (2011) define transferability as the degree to which a method is capable of providing comparable results, with minimal adaptions for different imaging conditions. For example, Kohli et al. (2013) used a fuzzy membership classifier and texture features combined with an adapted Global Slum Ontology (GSO) to map dwellings in India. The analysis was performed on three test sites and showed accuracies ranging from 68% to 47% with minimal adaptions. Belgiu et al. (2014) confirms the results of Kohli et al. (2013) and the results in this study by noting that different scene extents introduces variations in the size of image objects and hence feature statistics of image objects may differ, resulting in a moderate decrease in classification accuracy. Thus given the anticipated and noted change in accuracy between test sites and scene extents, the best potential and most viable informal settlement solution would be to run independent feature selections per test site, rather than seeking transferability in its methods.

4.6 Conclusion

This study tested the utility of high resolution WV-2 and aerial imagery within an OBIA environment for the objective identification of image segmentation, feature selection and classification parameters in the context of informal settlement mapping. An objective identification approach was important, as the selection of parameters is usually user-defined, error-prone, and the interpretation is biased by human subjectivity. The supervised segmentation parameter tuner (SPT) tool was used to reduce the trial-by-error process by deriving segmentation parameters for individual informal settlements. The SPT tool successfully derived optimal segmentation parameters for all test sites with high compactness and correctness values. The number of available object features for classification makes a detailed qualitative exploratory analysis of every individual image feature extremely time-and-effort intensive, which led to the introduction of feature selection methods. The RF algorithm were used to reduce data dimensionality of the WV-2 (n=140) and aerial imagery (n=60) feature sets by 23% and 53%, whereas the CART algorithm reduced the same feature set by 95% and 91% respectively. Nevertheless although different feature subsets were selected that mainly comprised of different features (geometry, mean layer, texture) similar classification accuracies were achieved in test sites 1, 2 and 3 for the features selected by the RF and CART algorithm, despite the test sites displaying varying informal settlement characteristics. This study was the first to include an nDSM to assist classification of informal dwellings within an informal settlement, and was significant as it received high importance scores from both the RF and CART algorithm using the WV-2 and aerial imagery.

CHAPTER 5: DISCUSSIONS AND CONCLUSIONS

A summary of the findings of the experiments in chapter 3 and 4 are presented in this chapter. The research aims and objectives are revisited and the research is critically assessed regarding the achievement of its stated objectives. The limitations of this work and opportunities for future studies are discussed. Finally, a concluding section is presented.

5.1 Summary of findings

Chapter 3 addressed and achieved **objective 1**, to investigated and extract nDSM models in a complex urban environment using: (i) LiDAR data, (ii) high resolution aerial photographs in a process of IM, and (iii) a series of aerial images captured using a hand-held camera using SfM techniques. Photogrammetric IM and LiDAR are traditional and well proven methods whereas SfM is a novel technique that has not been widely used for nDSM extraction. This study represented a first attempt at evaluating the three approaches together, particularly for mapping informal dwellings. Area-based metrics such as completeness and correctness was calculated to evaluate how well the extracted nDSM compared against accurate reference extents. Positional-based accuracy was calculated to assess the ability of the nDSM to separate objects above the DTM. Vertical profiles were calculated to evaluate the z-values of individual dwellings against a field reference. When assessing the accuracy of the models across the informal settlements, the accuracy of the models differed depending on the structural arrangements of dwellings. It was found that general increase in accuracy in the less sparse informal settlement. LiDAR produced the highest overall accuracy and would be used as ancillary dataset for classification in Chapter 4. The changing accuracy between methods highlighted the importance of selecting the most relevant method and/or dataset to model the type of informal settlements

Objective 2 was achieved in Chapter 4 using a systematic approach to objectively identify segmentation parameters, reduce feature dimensionality and assist classification. The SPT tool was used to optimally derive segmentation parameters for each informal settlement, as dwelling characteristics was too diverse to apply the segmentation parameter set from test site1 to test site 2 and 3. The aerial imagery was used as input to the SPT tool and derived representative dwelling extents. The accuracy of the segmentation was evaluated and achieved high completeness and correctness in all areas. The RF and CART algorithm were used to remove redundant and unnecessary features. Using the RF algorithm, the original WV-2 feature set

(n=140), aerial feature set (n=60) and combined WV-2 and aerial feature sets (n=200) was reduced to a subset of 108, 28 and 138 features, which mostly comprised of different features in the top 10, 20, 30 and feature subsets. Likewise the CART algorithm reduced the WV-2 features and aerial features to a subset of 6 and 5 features, with high importance given to the nDSM in both subsets. It was found that although different feature subsets were selected, similar classification accuracies were achieved in test sites 1, 2 and 3 for the features selected by the RF and CART algorithm. This revealed that there is no specific or best suited feature(s) for mapping informal dwellings, as a simple WV-2 or aerial feature subset selected by CART can achieve similar classification accuracies to a complex RF WV-2 or aerial feature subsets, and further demonstrates the challenge of mapping informal settlements. The results of this study is promising given that objective identification approaches were used to determine important segmentation, feature subsets selection and classification parameters rather than using a subjective trial- by-error approach.

5.1.1 Limitations of the study

In chapter 3, the results of the nDSM would be more comparable if both SfM and IM was extracted using the same dataset with regards to the same flying height, lens distortions, sun angle, field of view etc. These changing factors would ensure the same image overlap and results in more comparable nDSMs.

In addition different DTMs were used to extract the nDSM between the LiDAR, IM and SfM datasets. For the LiDAR nDSM a LIDAR derived DTM was used, whereas the SUDEM was used for the IM and SfM imagery, as no DTM was derived from the software used to create the IM DSM (Orthoengine) or the SfM DSM (Designing Reality). The different DTM used may account for the change in accuracy in test site 1 and test site 2 as the SUDEM is a 5m DTM and the LiDAR-derived DTM is 1m.

Significant uncertainty exists with image/data acquisition and processing for IM, SfM and LiDAR due to structural complexity, diversity and high degree of detail within informal settlements (Stal et al. 2013). It was noted that careful consideration needs to be given when selecting a preferred technique or dataset, as selection is a balance between desired accuracy, the cost involved in creation (acquisition and processing), availability of data and the projects constraints and requirements as limitations in model accuracy will propagate through to model predictions (Fryer et al. 1994).

In chapter 4, the supervised SPT tool uses user defined segments to optimally derive segmentation parameters. Since data did not exist, dwellings extents were manually digitized and may affect the outcome of SPTs area-based accuracy assessment.

5.1.2 Revisiting the research problem

The inclusion of an nDSM as an ancillary classification dataset addressed a recommendation by Pinho et al. (2011) for finding a solution to discriminate between bare soil and (oxidized) metallic roof structures. Furthermore the nDSM eliminated the need for multiclass classification as highlighted in Ballim, Poona & Ismail (2014) and Hoffman (2008), as dwellings were classed according to height about the ground (DTM) and not by inter-dwelling variations. It was found that the introduction of an objective identification approach for segmentation (SPT), feature selection and classification, reduced the human subjectivity and eliminated a time-consuming and trial-by-error process as previous described in literature. The objective identification approach was valuable in finding important parameters that was able to address the relatively complex and undefined morphology of informal settlements.

The results indicate that the WV-2 and aerial imagery was a suitable dataset, achieving similar classification results using the RF and CART feature subsets in test sites 1, 2 and 3. Furthermore the aerial imagery proved to be a valuable data source, assisting in dwelling field verifications, providing the input imagery for the SPT tool and as basis for the accuracy assessments.

Remote sensing and image processing methods allows up-to-date geospatial information for planning, monitoring and visualization of informal settlements to be obtained. Unlike survey, census, and participatory–based approaches that is limited by the accessibility of dwellings, remote sensing methods offers an alternative solution for methodology development that can provide predictable and consistent results for the identification of informal settlements using satellite and aerial imagery (Pinho et al. 2011).

By mapping informal settlements it will be possible to identify preferred locations that will allow upgrading efforts to be concentrated on older, more established settlements, where the chances of achieving successful in-situ rearrangement and upgrading are higher (Informal Settlements Handbook 2009). This can facilitate intervention based programs such as ‘re-blocking policy’ and ‘Upgrading informal settlement programme’ (UISP), as well as the implementation of the ‘Informal Settlement Atlas’. Similarly mapping can facilitate the National Department of Human

Settlements (NDH) in its implementation of the Upgrading informal settlement programme (UISP) with the objective to upgrade all informal settlements in the country. The findings from this research will contribute to the body of knowledge on remote sensing of human settlements and advanced image processing techniques to assist and/or improve the current method of enumeration, by providing an alternative methodology that is more efficient and accurate for managing informal dwellings and settlements.

5.2 Recommendations for future research

Possibilities for improving informal dwelling classification include deriving a more accurate nDSM using high resolution imagery for IM, SfM and denser datasets for LiDAR. Higher detailed elevation models will facillitates improved image classification (Sori 2011), by more accurately representing dwelling extents (Remondino et al. 2014). An investigation into novel systematic approaches to objectively identify segmentation and classification parameters can ensure robust informal settlement models devoid of human subjectivity. The transferability of the developed methodology can be further evaluated, by testing the segmentation and classification, to a wider spatial region that encompass more spatially and spectrally variable informal settlements. A comparison can be conducted for time and resources taken to compute remote sensing-based dwellings extents (including DEM generation, segmentation, feature selection and classification) versus census and participatory approaches. Alternatively, remote sensing-based methods can assist census and participatory approaches by providing useful information

5.3 Concluding remarks

The aims of this study were duly accomplished. This study represents the first attempt at identifying individual dwellings within an informal settlement using a semi-automated objective approach for segmentation, feature selection and classification. This study is also the first to evaluate different approaches to derive nDSMs in an informal settlement using traditional and well understood techniques/data such as LiDAR and IM against a novel technique such as SfM. The robustness and transferability of the models was evaluated by application to independent test sites, in order to demonstrate the capability of our model to provide comparable results.

The results of this research show great promise as an efficient approach to mapping and monitoring informal dwellings within an informal settlement. The relatively low number of international publications on OBIA-based informal settlement mapping , confirms the challenges for accurate methodology development, as the dynamic spatial-temporal behaviour, the relatively high inner-structural heterogeneity (Hoffman et al. 2008, Pinho et al. 2011, Shekhar 2012) and

microstructure of informal settlements do not allow for generically applicable urban mapping solutions to be produced (Taubenbock and Kraff 2013). This study however represents an attempt to address the relatively complex and undefined morphology of informal settlements by combing remote sensing (RS) and advanced image processing methods for methodology development that is both robust and transferable.

Classification of informal dwellings using RS methods can provide predictable and consistent results that will allow alternative housing options to be strategically planned, such as the development of implementation programs and the evaluation of in-situ upgrading progress. More specifically the classification of informal dwellings can assist in addressing Stats SA Census under-coverage associated with enumeration of dwellings in informal settlements, and the lack of informal dwelling representation in Eskom's Spot Building Count. Additionally the methodology provides a fast and accurate method for updating and monitoring the Informal Settlement Atlases that was compiled by the CSIR, AfriGIS, and Western Cape Department of Human Settlements.

REFERENCES

Addink EA, De Jong SM, Davis SA, Dubyanskiy V, Burdelov LA & Leirs H 2010. The use of high-resolution remote sensing for plague surveillance in Kazakhstan. *Remote Sensing of Environment* 114: 674-681.

Allen K & Heese K 2013. Understanding why service delivery protests take place and who is to blame. Available from: http://www.municipaliq.co.za/[Accessed September 2015].

Alliance 2015. Millennium Development Goals. Available from http://www.alliance2015.org/fileadmin/user_upload/MDGs.pdf [Accessed August 2015].

Agarwal S, Snavely N, Simon I, Seitz SM, Szeliski R 2010 *Building Rome in a day.* Proceedings of the IEEE 12th International Conference on Computer Vision: 72–79. 29 September–2 October 2009.

AgiSoft. Agisoft PhotoScan; Available online: http://www.agisoft.ru/products/photoscan/ (accessed on 20 October 2015).

Aguilar MA, Saldaña MDM, & Aguilar FJ 2012. Generation and quality assessment of stereo-extracted DSM from GeoEye-1 and WorldView-2 imagery. *Geoscience and Remote Sensing, IEEE Transactions* 52 (2): 1259–1271.

Aminipouri M 2009.Object oriented Analysis of very high resolution orthophotos for estimating the population of slum areas, case of Dar-es-Salam, Tanzania. Master's thesis. International Institute for Geo-Information Science and Erath Observation. Enschede, The Netherlands.

Andrews DP, Bedford J & Bryan PG 2013. *A comparison of laser scanning and structure from motion as applied to the great barn at harmondsworth,UK*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-5/W2. XXIV International CIPA Symposium, 2–6 September 2013, Strasbourg, France.

Arvor D, Durieux L, Andrés S & Laporte MA 2013. Advances in geographic object-based image analysis with ontologies: a review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing* 82: 125–137.

Awrangjeb M & Fraser C 2012. Rule-based segmentation of lidar point cloud for automatic extraction of building roof planes. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W3 CMRT13-City Models, Roads and Traffic, 12–13 November 2013, Antalya, Turkey.

Baatz M & Schäpe A 2000. *Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation*. J. Strobl, T Blaschke, G Griesebner (Eds.), Angewandte Geographische Informations-Verarbeitung XII, Wichmann Verlag, Karlsruhe, Germany: 12–23.

Ballim K, Poona NK & Ismail 2014. Mapping Informal Dwellings from WorldView-2 data in Cape Town using an Object-Based Image Analysis approach. *South-Eastern European Journal of Earth Observation and Geomatics* 3 (2S): 299-303.

Benz UC, Hofmann P, Willhauck G, Lingenfelder I & Heynen M 2004. Multiresolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3–4): 239–258.

Belgui M, Drăgut L & Strobl J 2013. Quantitative evaluation of variations in rule-based classifications of land cover in urban neighbourhoods using WorldView-2 imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 85: 205-215.

Belgui M & Drăgut L 2016. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114: 24-31

Blaschke T, Lang S & Hay G 2008. Object Based Image Analysis. Springer. Heidelberg, Berlin, New York: 817.

Blaschke T 2010. Object Based Image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 65: 2-16.

Briem GJ, Benediktsson JA & Sveinsson JR 2002. Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing* 40: 2291– 2299.

Breiman L 2001.Random forests. *Machine learning* 45: 5–32.

Breiman L 1998. Randomisation outputs to increase prediction accuracy. Technical report, Statistics department, University of California, Berkeley,

Breima L, Friedman JH, Olshen RA & StoneCJ 1984.Classification and regression trees. Belmont: Wadsworth.

Brédif M, Tournaire O, Vallet B & Champion N 2013. Extracting polygonal building footprints from digital surface models: a fully-automatic global optimization framework. *ISPRS Journal of Photogrammetry and Remote Sensing* 77: 57–65.

Boufama B,Mohr R. & Veillon F 1993. *Euclidean constraints on uncalibrated reconstruction*. Proceedings of the Fourth International Conference on Computer Vision, Berlin, Germany: 466–470.

Bolles RC & Fischler MA 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24(6): 381–395.

Boufama B, Mohr R. & Veillon F 1993. *Euclidean constraints on uncalibrated reconstruction.* Proceedings of the fourth international conference on computer vision, Berlin, Germany: 466–470.

Böhler W, Bordas V & Marbs A 2003. Investigating Laser Scanner Accuracy. *International Archives of Photogrammetry and Remote Sensing* 34: 696–701.

Buján S, González- Ferreiro E, Barreiro- Fernández L, Santé I, Corbelle E & Miranda D 2013. Classification of rural landscapes from low-density lidar data: is it theoretically possible?, *International Journal of Remote Sensing* 34:16.

Burrough PA, McDonnell RA & Lloyd CD 2015. Principles of geographical information systems. Oxford University Press.

Campbell JB 2006. Introduction to Remote Sensing. 4th ed. Abingdon: Taylor & Francis.

Carleer A, Debeir O & Wolff E, 2005. Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing* 71 (11): 1285–1294.

Ceylan D, Mitra NJ, Zheng Y, Pauly M 2014. Coupled structure-from-motion and 3D symmetry detection for urban facades. *ACM Transactions on Graphics* 33(2): 1–15.

Champion N, Matikainen L, Rottensteiner F, Liang X & Hyyppä J 2008. *A test of 2D building change detection methods: comparison, evaluation and perspectives*. Proceedings of the XXIst ISPRS Congress, Beijing, China, 3–11 July 2008. XXXVII (B4): 297-303.

Chan JCW, Beckers P, Spanhove T & Borre JV 2012. An evaluation of ensemble classifiers for mapping Natura 2000 heathland in Belgium using spaceborne angular hyperspectral (CHRIS/Proba) imagery. *International Journal of Applied Earth Observation* 18: 13–22.

Chen C, Wakefield J, Lumely T 2014. The use of sampling weights in Bayesian hierarchical models for small area estimation. *Spatial and Spatio-temporal Epidemiology* (11): 33-43.

Cheuk ML & Yuan M 2009. Assessing spatial uncertainty of LiDAR-derived building model: A case study in downtown Oklahoma city. *Photogrammetric Engineering & Remote Sensing* 74 (12): 257–269.

Childs C 2004. Interpolating surfaces in ArcGIS spatial analyst. ArcUser, July-September.

Church E, Quinn A 1948. Elements of Photogrammetry. Syracuse, NY, Syracuse University Press

Chubey Ms, Franklin SE & Wulder MA 2006. Object-based analysis of Ikonos-2 imagery for extraction of forest inventory parameters. *Photogrammetric Engineering and Remote Sensing* 72: 383-394.

Cornelis N, Leibe B, Cornelis K & Van Gool L 2008. 3D urban scene modeling integrating recognition and reconstruction. *International Journal of Computer Vision* 78(2-3): 121-141.

Congalton RG, Green K & 2009. Assessing the accuracy of remotely sensed data: principles and practices 2nd ed. CRC Press, Boca Raton.

CUDA 2013. CUDA Programming Guide. NVidia. [Online Help] Online available: http://docs.nvidia.com/cuda/cuda-c-programming-guide/index.html.

Davis L 1991. Handbook of generic algorithms. Van Nostrand Reinhold.

De'ath G and. Fabricius KE 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.

Demir N, Poli D & Baltsavias E 2009. Detection of buildings at airport sites using images & LiDAR data and a combination of various methods. *Archives of Photogrammetry,Remote Sensing and* SIS 38: 71-77.

DLR (Deutsches Zentrum für Luft- und Raumfahrt) 2013. [Online] available from: http://www.stadtentwicklung.berlin.de/umweltatlas/download/e0610_Edition2013_Documentation_ Building_Vegetation_Heights.pdf [Accessed March 2015].

Dronova I 2015. Object-Based Image Analysis in Wetland Research: A Review. *Journal of Remote Sensing* 7: 6380-6413.

Dragut L, Csillik O, Eisank C & Tiede D 2014. Automated parameterisation for multi-scale image segmentation on multiple layers. *ISPRS Journal of Photogrammetry and Remote Sensing* 88: 119–127.

Drăguţ L &Eisank C 2011. Object representations at multiple scales from digital elevation models. *Geomorphology* 129: 183–189.

Dragut L, Tiede D, Levick S 2010. ESP: a tool to estimate scale parameters for multiresolution image segmentation of remotely sensed data. *International Journal of Geographical Information Science* 24: 859– 871.

Du S, Zhang F & Zhang X 2015. Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. *ISPRS journal of photogrammetry and remote sensing* 105: 107–119.

Duro DC, Franklin SE, Dub MG, 2012. A comparison of pixel-based and object based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment* 118: 259–272.

eCognition 2012. eCognition Developer 8.8 Reference Book. Trimble Germany.

Ekhtari N, Sahebi MR, Valadan Z & Mohammadzadeh MJ 2008. *Automatic Building Detection from LIDAR Point Cloud Data*. Proceedings of the XXIst ISPRS Congress, Beijing, China, 3–11 July 2008; XXXVII (B4): 473-477.

Eos Systems Inc. PhotoModeler; Available online: http://www.photomodeler.com/products/photomodeler.htm / (accessed on 20 October 2015).

Feitosa RQ, Costa GAOP, Cazes TB & Feijo B 2008. Genetic adaption of segmentation parameters. In: Blaschke T, Hay G, Lang S (Eds.) Object-Based Image Analysis–Spatial Concepts for Knowledge-Driven Remote Sensing Applications. Lecture Notes in Geoinformation & Cartography 18, Berlin, Springer: 679–695.

Flood M 2002. Product definitions and guidelines in specifying lidar deliverables, *Photogrammetric Engineering & Remote Sensing* 68(12): 1230–1234.

Freund Y & Schapire RE 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer Systems and Science 55*: 119–139.

Fryer JG, Chandler JH & Cooper MAR 1994. On the accuracy of heighting from aerial photographs and maps: implications to process modellers. *Earth Surface Processes and Landforms* 19: 577-583.

Fonstad MA, Dietrich JT, Courville BC, Jensen LJ & Carbonneau PE 2013. Topographic Structure from Motion: a new Photogrammetric Measurement. Earth Surface Processes and Landform 38: 421-430.

Furukawa Y & Ponce J. 2010. Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32: 1362–1376.

McGlone JC, Lee GY 2013. Manual of photogrammetry.6th ed. American Society for Photogrammetry and Remote Sensing: Bethesda, Md.

Gao J 2009. Digital Analysis of Remotely Sensed Imagery. New York: McGraw Hill.

Gao Y, Mas JF, Kerle N, Navarrete Pacheco JA 2011. Optimal region growing segmentation and its effect on classification accuracy. *International Journal of Remote Sensing* 32(13): 3747–3763.

Garnero G & Godone D 2013. *Comparisons between different interpolation techniques*. Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-5 W (3): 27-28.

Geomatica Orthoengine User Guide 2003. PCI Geomatica. [Online] available from: https://www.ucalgary.ca/appinst/doc/geomatica_v91/manuals/orthoeng.pdf [Accessed March 2014].

Gehrke S, Morin K, Downey M, Boehrer N & Fuchs T 2010.Semi-global matching: an alternative to LiDAR for DSM generation?. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38.

Ghose MK, Pradhan R & Ghose SS 2010. Decision Tree Classification of Remotely Sensed Satellite Data using Spectral Separability Matrix. *International Journal of Advanced Computer Science and Applications* 1(5).

Ghosh A, Fassnacht FE, Joshi PK & Koch B 2014. A framework for mapping tree species combining hyperspectral and LiDAR data: role of selected classifiers and sensor across three spatial scales. *International Journal of Applied Earth Observation* 26: 49–63.

Gruen A 1985. Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography* 14(3): 175–187.

Golparvar-Fard M, Pena-More F & Savarese S 2011. *Monitoring changes of 3D building elements from unordered photo collections*. Proceedings of the computer vision workshop held 6-13 November 2011. Barcelona: IEEE international conference.

Gruen A 2012.Development and Status of Image Matching in Photogrammetry. *The Photogrammetric Record* 27 (137): 36–57.

Habitat Agenda 1996. Available from: http://www.un-documents.net/ha-1.htm. [Accessed November 2015].

Hay GJ, Castilla G, Wulder MA & Ruiz JR 2005. An automated object-based approach for the multiscale image segmentation of forest scenes. *International Journal of Applied Earth Observation* 7: 339-359.

Hamedianfar A, Shafri HZM, Mansor S & Ahmad N 2014. Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and LiDAR data, *International Journal of Remote Sensing* 35(5): 1876-1899.

Haralick RM., Shanmugam K, Dinstein I 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3 (6): 610–621.

Harika B, Rao NR & Krishnaiah RV 2013. A novel tool for stereo matching of images. *International Journal of Research in Engineering and Technology* 2(10): 323-327.

Hermosilla T. Ruiz LA, Jorge AR & Estronell J 2011. Evaluation of automatic building detection approaches combining high resolution images and lidar data. *Journal of Remote Sensing* 3: 1188-1210.

Hamedianfar A, Shafri HZM, Mansor S & Ahmad N (2014) Improving detailed rule-based feature extraction of urban areas from WorldView-2 image and lidar data. *International Journal of Remote Sensing* 35(5): 1876-1899.

Hapfelmeier A & Ulm K 2013. A new variable selection approach using random forests. *Journal of computational statistics and data analysis* 60: 50–69.

Hirschmuller H, Buder M & Ernst I 2012. *Memory efficient semi-global matching*. Proceedings of the 22nd ISPRS Congress held 25 August - 1 September 2012. Australia. Melbourne: International Society for Photogrammetry and Remote Sensing.

Hofmann P, Strobl J, Blaschke T & Kux H 2008. Detecting Informal Settlements from QuickbirdData in Rio De Janeiro Using An Object Based Approach. In Object-Based Image Analysis; Springer: 531–553.

Hofmann P 2001. *Detecting informal settlements from Ikonos image data using methods of object oriented image analysis-an example from Cape Town (South Africa)*. Proceedings of remote sensing of urban areas conference, Regensburg, Germany: 107-118.

Huising EJ & Pereira LMG 1998. Errors and accuracy estimates of laser data acquired by various laser scanning systems for topographic applications, *ISPRS Journal of Photogrammetry & Remote Sensing* 53: 245–261.

Hughes G 1968. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory* 14: 55-63.

Im J, Lu Z, Rhee J & Quackenbush LJ2012. Impervious surface quantification using a synthesis of artificial immune networks and decision/regression trees from multi-sensor data. *Remote Sensing of Environment* 117: 102–113.

Informal Settlements Handbook. South Africa (Republic of) 2009.

Ismail R and Mutanga O 2010. A comparison of regression tree ensembles: Predicting sirex noctilio induced water stress in Pinus patula forests of KwaZulu-Natal, South Africa. *International Journal of Applied Earth Observations and Geoinformation* 125: 45–51.

Janak RJ 2010. Improving the Quality of Digital Surface model generated from very high

resolution satellite stereo imagery by using object oriented image analysis technique. Master's thesis. International Institute for Geo-information Science and Earth Observation. Enschede,Netherlands.

Jancosek M & Pajdla T 2011. *Multi-View Reconstruction Preserving Weakly-Supported Surfaces (CVPR)*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition held 20-25 June 2011. Providence. Rhode Island

Jain A & Zongker D 1997. Feature selection: evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(2):153–158.

Jensen JR 2007. *Remote Sensing of the Environment: An Earth Resource Perspective*. 2nd edition. Prentice Hall.

Johnson B & Xie Z 2011. Unsupervised image segmentation evaluation and refinement using a multi-scale approach. *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (4): 473–483.

Joglekar J, Gedam SS 2012. Area based image matching methods—A survey. *International Journal of Emerging Technology and Advanced Engineering* 2(1): 130-136.

Kersten T, Mechelke K & Maziull L 2015. 3d model of al Zubarah fortress in Qatar – terrestrial laser scanning vs. Dense image matching. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5/W4, 2015 3D Virtual Reconstruction and Visualization of Complex Architectures, 25-27 February 2015, Avila, Spain.

Kit O, Lüdeke M 2013. Automated detection of slum area change in Hyderabad, India using multitemporal satellite imagery. ISPRS Journal of Photogrammetry and Remote Sensing 83: 130-137.

Kohli D, Warwadekar P, Kerle N, Sliuzas R & Stein A 2013 .Transferability of object-oriented image analysis methods for slum identification. Journal of Remote Sensing 5: 4209-4228.

Kohli D, Sliuzas R, Kerle N & Stein A 2012. An ontology of slums for image-based classification. Journal of Computers, Environment and Urban Systems 36: 154–163.

Kohavi R & John GH 1997. Wrappers for feature subset selection *Artificial intelligence* 97: 273–324.

Krauβ T & d'AngeloP 2011. Morphological filling of digital elevation models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXVIII-4/W19, ISPRS Workshop held at Hannover 2011, 14-17 June 2011, Hannover, Germany.

Kursa MB & Rudnicki 2011. *A deceiving charm of feature selection:* The microarray case study, in Man-Machine Interactions 2, vol. 103.T Czachorski, S Kozielski & UStanczyk. Eds. Berlin, Germany: Springer-Verlag, 2011: 145–152.

Kursa MB 2012. Important attribute search using Boruta algorithm, in: Package "Boruta". Available from: http://cran.r-project.org/web/ packages/randomForest/index.html

Kursa MB & Rudnicki WR 2010. Feature selection with the Boruta package. *Journal of Statistical Software* 36: 1–13.

Kursa MB, Jankowski A, and Rudnicki WR 2010. Boruta-A system for feature selection, *Fundamenta Informaticae* 101: 271–285.

Lawrence RL, Wood SD & Sheley RL 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of the Environment* 100*: 356–362.*

Laliberte AS, Fredrickson EL & Rango A 2007. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogrammetric Engineering & Remote Sensing* 73: 197-207.

Laliberte AS, Browning DM & Rango A 2010. Feature selection methods for object-based classification of sub-decimeter digital aerial imagery. *The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences*: GEOBIA 2010, 29 Jun – July 2, 2010, Ghent, Belgium, ISPRS, Vol. No. XXXVIII-4/C7.

Laliberte AS, Browning DM & Rango A 2012. A comparison of three feature selection methods for object-based classification of sub-decimeter resolution UltraCam-L imagery. *International Journal of Applied Earth Observation and Geoinformation* 15: 70–78.

Leberl F, Meixner P & Brédif M 2011. *3D roof details by 3D aerial vision*. Proceedings of IEEE International Conference on Computer Vision Workshops: 212-218.

Lee JS, Grunes MR & Pottier E 2001. Quantitative comparison of classification capability: fully polarimetric versus dual and single polarization SAR. *IEEE Transactions on Geoscience and Remote Sensing* 39(11): 2343-2351.

Liaw A and Wiener M 2002. Classification and regression by random Forest. R News 2: 18–22.

Lee JS, Grunes MR & Pottier E 2001. Quantitative comparison of classification capability: fully polarimetric versus dual and single polarization SAR. *IEEE Transactions on Geoscience and Remote Sensing* 39(11): 2343-2351.

Li M & Liu Y 2015. Training set size, scale, and features in Geographic Object-Based Image Analysis of very high resolution unmanned aerial vehicle imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 102: 14–27.

Li X, Myint SW, Zhang Y, Galletti C, Zhang X & Turner BL 2014. Object-based land-cover classification for metropolitan Phoenix, Arizona, using aerial photography. *International Journal of Applied Earth Observation and Geoinformation* 33: 321-330.

Liu D & Xia F 2010. Assessing object-based classifications: advantages and limitations. *Remote Sensing Letters* 1(4): 187-194.

Lowe DG 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60(2): 91-110.

MacFaden SW, O'Neil-Dunne JPM, Royar AR, Lu JWT & Rundle AG 2012. High- Resolution Tree Canopy Mapping for New York City Using LIDAR and Object-Based Image Analysis. *Journal of Applied Remote Sensing* 6 (1).

Maiellaro N, Zonno M & Lavalle P 2015. *Laser scanner and camera-equipped uav architectural surveys*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5/W4, 2015 3D Virtual Reconstruction and Visualization of Complex Architectures, 25-27 February 2015, Avila, Spain.

Maimon, O & Rokach L. 2005 (Eds.). The Data Mining and Knowledge Discovery Handbook :164-166. Springer US.

Mallet C & Bretar F 2009. Full-waveform topographic Lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing* 64(1): 1-16.

Mallet C, Bretar F, Roux M & Soerge U 2011. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS journal of photogrammetry and remote sensing* 66: 571–584.

Mashimbye ZE, de Clercq PW & Van Niekerk A 2013. An evaluation of digital elevation models (DEMs) for delineating land components. *Geoderma* 213: 312- 319.

Maiellaro N, Zonno M & Lavalle P 2015. Laser scanner and camera-equipped uav architectural surveys. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-5/W4: 381-386.

Martha TR, Kerle N, van Westen CJ, Jetten V & Kumar KV, 2011. Segment optimization and data-driven thresholding for knowledge-based landslide detection by object-based image analysis. *IEEE Transactions on Geoscience and Remote Sensing* 49: 4928–4943.

Martina HB & Ginzler C 2012. Accuracy assessment of digital surface models based on WorldView-2 and ADS80 stereo remote sensing data. *Sensors Journal* 12(5): 6347–6368.

Mathenge CW 2011. Application of object oriented image analysis in slum identification and mapping –the case of Kisumu, Kenya. Master's thesis. University of Twente, Department of Geo-information Science and Earth Observation.

Mathews AJ & Jensen JLR 2014. Three-dimensional building modelling using structure from motion: improving model results with telescopic pole aerial photography. Papers of the Applied Geography Conferences 35: 98–107.

McNemar Q 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 3 (2): 153-157.

Miao X, Heaton JS, Zheng S, Charlet DA & Liu H 2012. Applying tree-based ensemble algorithms to the classification of ecological zones using multitemporal multi-source remote-sensing data. *International Journal of Remote Sensing* 33: 1823–1849.

Micheletti N, Chandler JH & Lane SN 2015. Geomorphological Techniques. British Society for Geomorphology [online], Chap 2, Sec 2.2. Available from: http://www.geomorphology.org.uk/sites/default/files/geom_tech_chapters/2.2.2_sfm.pdf [Accessed 26 February 2016].

Microsoft Corporation. Photosynth; Available online: http://photosynth.net/ (accessed on 20 October 2015).

Mie J, Dawei HE, Habetler T & Guannan QU 2014. *A random forest method for real-time price forecasting in New York electricity market.* IEEE Conference & Exposition, PES General Meeting held 27-31 July 2014, National Harbor, MD.

Millard K. & Richardson M 2015. On the importance of training data sample selection in random forest image classification: a case study in peatland ecosystem mapping. *Remote Sensing* 7: 8489-8515.

Millenium Project. Available online: http://www.unmillenniumproject.org [Accessed on February 2014].

Mikolajczyk K & Scdmid C 2002. *An affine invariant interest point detector*. Proceeding of the 7th European conference on Computer Vision held on 28-31 May 2002.

Moreira MJM, Nex F, Agugiaro G, Remondino F & Lim NJ 2013. *From dsm to 3d building models: a quantitative evaluation*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,Volume XL-1/W1, ISPRS Workshop held at Hannover 2013, 21–24 May 2013, Hannover, Germany.

Myint SW, Gober P, Brazel A, Grossman-Clarke S &Weng Q 2011. Perpixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. *Remote Sensing of Environment* 115:1145–1161.

Navulur K 2007. Multispectral image analysis using the object-oriented paradigm. CRC Press, Boca Raton, FL.

Nilsson R, Pena J, Bjorkegren J, Tegner J 2007. Consistent Feature Selection for Pattern Recognition in Polynomial Time. *The Journal of Machine Learning Research* 8: 612.

Novack T, Esch T, Kux H , Stille U 2011. Machine Learning Comparison between WorldView-2 and QuickBird-2-Simulated Imagery Regarding Object-Based Urban Land Cover Classification. *Journal of remote sensing* 3: 2263–2282.

Nussbauma S, Niemeyerb I & Cantya MJ 2008. *Seath -a new tool for automated feature extraction in the context of object-based image analysis*. ISPRS Conference proceedings. XXXVI- 4: C42.

Olsen B 2004. Automatic Change Detection for Validation of Digital Map Databases. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXV (B2): 569-574.

O'Neil-Dunne JPM, MacFaden SW, Royar AR & Pelletier KC 2012. An Object-Based System for LiDAR Data Fusion and Feature Extraction. *Geocarto International* 1–16.

Pal Laliberte AS, Browning DM & Rango A 2010. Feature selection methods for object-based classification of sub-decimeter digital aerial imagery. *The International Archives of the Photogrammetry, Remote Sensing, and Spatial Information Sciences*: GEOBIA 2010, 29 Jun – July 2, 2010, Ghent, Belgium, ISPRS, Vol. No. XXXVIII-4/C7.

Pal M & Foody GM 2010. Feature selection for classification of hyperspectral data by SVm. *IEEE Transactions on Geoscience and Remote Sensing* 48: 2297-2307.

Pinho CM, Fonseca LMG, Korting TS, Almeida CM & Kux HJH 2011. Land-cover classification of an intra-urban environment using high-resolution images and object-based image analysis, *International Journal of Remote Sensing* 33 (19): 5973-5995.

Poli D, Remondino F, Angiuli E & Aguigior G 2015. Radiometric and geometric evaluation of GeoEye-1, WorldView-2 and Pléiades-1A stereo images for 3D information extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 100: 35-47.

Pooja AP, Jayanth J & Koliwad S 2011. Classification of RS data using Decision Tree Approach. *International Journal of Computer Applications* 23 (3).

Poona NK & Ismail R 2014. Using Boruta-Selected Spectroscopic Wavebands for the Asymptomatic Detection of Fusarium Circinatum Stress. *IEEE journal of selected topics in applied earth observations and remote* sensing 7(9): 3764-3772.

Poon J, Fraser CS, Zhang C 2007.Digital surface models from high resolution satellite imagery," *Photogrammetric Engineering & Remote Sensing* 73(11): 1225–1232.

Pudil P, Novovicova J & Kittler J 1994. Floating search methods in feature selection. *Pattern Recognition Letters* 15(11): 1119–1125.

Qin R 2014. Change detection on LOD 2 building models with very high resolution spaceborne stereo imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 96: 179-192.

R Development Core Team. (2011). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing [Online]. Available: http://www.r-project.org/ [Accessed: June 2015].

Rahman MR & Saha SK 2008. Multiresolution Segmentation for Object-based Classification and Accuracy Assessment for Land use/Land Cover Classification using Remotely Sensed Data. Journal of the Indian Society of Remote Sensing 36 (2).

Rejaur R and Saha S 2008. Multi-resolution segmentation for object-based classification and accuracy assessment of the land use / land cover classification using remote sensed data. *Journal of the Indian Society of Remote Sensing* 36: 189-201.

Remondino F, Spera MG, Nocerino E, Menna F,Nex F 2014. State of the art in high density image matching. *The Photogrammetric Record* 29: 144–166.

Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M & Rigol-Sanchez JP 2012. An assessment of the effectiveness of a random forest classifier for land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 67: 93-104.

Rutzinger M ,Rottensteiner F & Pfeifer N 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2(1): 11–20.

Salehi B, Zhang Y, Zhong M 2011. *Object-Based Land Cover Classification of Urban Areas Using VHR Imagery and Photogrammetrically-Derived DSM*. Proceedings of the ASPRS 2011 Annual Conference. Milwaukee, Wisconsin. May 1-5 2011.

Saha K, Wells NA & Munro-Stasiuk M 2011. An object-oriented approach to automated landform mapping: A case study of drumlins. *Computers and Geosciences* 37: 1324-1336.

Schapire RE 1990. The strength of weak learnability. *Machine Learning* 5: 197–227.

Shang J, Holmtrom D, McNairn H, Champagne C & Reicbert G 2009. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories. ISPRS Journal of Photogrammetry and Remote Sensing 64: 434-449.

Schenk T 2004. From point-based to feature-based aerial triangulation. *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (5): 315-329.

Schenk T 2005. Introduction to photogrammetry. The Ohio State University, Columbus.

Sibson R 1981. A Brief Description of Natural Neighbour Interpolation, Interpolating multivariate data: 21-36. John Wiley & Sons, New York.

Spetsakis ME, Aloimonos Y 1991. A multi-frame approach to visual motion perception. *International Journal of Computer Vision* 6: 245–255.

Salfords Systems 2001. CART- Tree-structured Non-Parametric Data Analysis [online]. Available from: http://docs.salford-systems.com/CART_Main.pdf [Accessed 2 August 2015].

Schaer P, Skaloud J. Landtwing S & Legat K 2007. *Accuracy Estimation for Laser Point Cloud Including Scanning Geometry*. Proceedings of The 5th International Symposium on Mobile Mapping Technology, Padua, Italy, 29–31 May 2007.

Schenk T 1999. *Digital Photogrammetry*: Volume I: 421. TerraScience, Laurelville, Ohio, USA.

Shekhar S 2012. Detecting slums from QuickBird data in Pune using an object-oriented approach. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 39 (8): 519-524.

Sebari I & He DC 2013. Automatic fuzzy object-based analysis of VHSR images for urban objects extraction. *ISPRS Journal of Photogrammetry and Remote Sensing* 79: 171-184.

Sliuzas R, Mboup G &de Sherbinin A 2008. Report of the expert group meeting on slum identification and mapping. Report by CIESIN, UN-Habitat ITC: 36.

South African SDI Alliance 2013 [online]. Available from: http://sasdialliance.org.za/ [Accessed January 2015].

Sori DN 2012. Identifying and classifying slum development stages from spatial data. Master's
thesis. University of Twente, Department of Geo-information Science and Earth Observation.

Spetsakis ME & Aloimonos Y 1991. A multi-frame approach to visualmotion perception. *International Journal of Computer Vision* 6: 245–255.

Stumpf A & Kerle N 2011. Object-oriented mapping of landslides using random forests. *Journal of Remote Sensing Environment* 115: 2564–2577.

Statistics South Africa 2011. Version 6. Census Undercount and strategies.

Su W Li J, Chen Y, Liu Z, Zhang J, Low TM, Suppiah I & Hashim SAM 2008. Textural and local spatial statistics for the object-oriented classification of urban areas using high resolution imagery. *International Journal of Remote Sensing,* 29 (11): 3105–3117.

Sun S & Salvaggio C 2013. Aerial 3D building detection and modelling from airborne LIDAR point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6: 1440–1449.

Snavely N, Seitz SM & Szeliski R. 2006. [online] Photo tourism. presented at the ACM SIGGRAPH 2006 Papers. Boston, Massachusetts. 835 pp. Available from: http://portal.acm.org/citation.cfm?doid=1179352.1141964 [Accessed August 2014].

Snavely N, Seitz SM & Szeliski R 2007. Modeling the World from Internet Photo Collections. *International Journal of Computer* Vision 80: 189–210.

Snavely KN 2008. Scene reconstruction and visualization from Internet photo collections, Ph.D., University of Washington.

Sibson R 1981. A Brief Description of Natural Neighbor Interpolation, Chapter 2 in Interpolating Multivariate Data, 21–36. John Wiley & Sons: New York.

Stasolla M & Gamba P 2007. Exploiting spatial patterns for informal settlement selection in arid environments using optical spaceborne data. Paper presented at photogrammetric image analysis, Munich, Germany.

Stal C, Tack F, De Maeyer P, De Wulf A & Goossens R 2013. Airborne photogrammetry and lidar for DSM extraction and 3D change detection over an urban area – a comparative study *International Journal of Remote Sensing* 34 (4): 1087-1110.

Tarantin E & Figorito B 2011. Extracting Buildings from True Color Stereo Aerial Images Using a Decision Making Strategy. *Journal of Remote sensing 3*: 1553-1567.

Taubenböck H & Kraff NJ 2013. The physical face of slums: a structural comparison of slums in Mumbai, India, based on remotely sensed data. *Journal of Housing and the Built Environment.*

Tiede D, Lang S, Hölbling D & Füreder P 2010. Transferability of obia rulesets for idp camp analysis in Darfur. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII-4/C7.

The Housing Development Agency 2012. Western Cape: Informal settlements status. Available from:http://www.thehda.co.za/uploads/files/HDA_Informal_settlements_status_Western_Cape.pdf [Accessed July 2014].

The Istanbul Declaration on Cities and Other Human Settlements 1996. Available from: http://www.dhs.gov.za/sites/default/files/legislation/Istanbul_Declaration.pdf [Accessed November 2015].

Trimble 2011. eCcognition Developer 8.64.1 Reference Book. Tribmle Germany GmbH, Munich.

Ullman S 1979. The Interpretation of Visual Motion. MIT Press: Cambridge, Mass.

UN 2003. The challenge of slums - global report on human settlements. United Nations Human Settlements Programme, Nairobi.

UN-Habitat 2012/2013.United Nations Human Settlement Programme. State of the World Cities: Prosperity of Cities.

UN-Habitat 2012/2013.United Nations Human Settlement Programme. State of the World Cities: Prosperity of Cities.

UN-Habitat 2006. State of the World's Cities: Trends in Sub-Saharan Africa, Urbanization & Metropolitanisation. New York: UN-Habitat.

Ussyshkin V & Theriault 2011. Airborne LiDAR: Advances in Discrete Return Technology for 3D Vegetation Mapping. *Journal of remote sensing* 3: 416 – 434.

Vancouver Declaration on Human Settlements 1976. Available from: http://www.dhs.gov.za/sites/default/files/legislation/Vancouver_Declaration_on_Human_Settlements_1976.pdf. [Accessed November 2015].

Van Niekerk A 2011. Stellenbosch University Digital Elevation Model Product Description, 2011 edition. Centre for Geographical Analysis, Stellenbosch.

Vu HH, Labatut P, Pons JP & Keriven R. 2012. High accuracy and visibility-consistent dense multiview stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(5): 889–901.

Wang Y, Chen S, & Cattani C 2012. Key Issues in Modelling of complex 3D Structures from Video Sequences. Mathematical Problem in Engineering 2012.

Western Cape Department of Environmental Affairs and Development Planning 2009. Western Cape Provincial Spatial Development Framework. Cape Town: Western Cape Department of Environmental Affairs and Development Planning.

Westoby M, Brasington J, Glasser NF & Hambrey MJ 2011. Structure-from-Motion: a high resolution, low-cost photogrammetric tool for geoscience applications. *Geomorphology* 179: 300–314.

Witharana C & Civco DL 2014. Optimizing multi-resolution segmentation scale using empirical methods: exploring the sensitivity of the supervised discrepancy measure Euclidean distance 2. *ISPRS Journal of Photogrammetry Remote Sensing* 87: 108–121.

World Health Organisation 2013. Global Health Observatory: Slum residence. Available from: http://www.who.int/gho/urban_health/determinants/slum_residence_text/en/ [Accessed June 2015].

Xu H 2013. Rule-based impervious surface mapping using high spatial resolution imagery. *International Journal of Remote Sensing* 34 (1): 27–44.

Zhang H, Fritts J & Goldman S 2008. Image segmentation evaluation: a survey of unsupervised methods. *Computer Vision and Image Understanding* 110 (2): 260–280.

Zhang YJ 1996. A survey on evaluation methods for image segmentation. *Pattern Recognition* 29 (8): 1335–1346.

Zhang C & Fraser C, 2009. An improved approach for DSM generation from high resolution satellite imagery. *Journal of Spatial Science* 54(2): 1-13.

Zhang J 2010. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion* 1(1): 5-24.

Zeng C, Wang J, Zhan W, Shi P & Gambles A 2014. An elevation difference model for building height extraction from stereoimage- derived DSMs. *International Journal of Remote Sensing* 35(22): 7614-7630.

Zhou W, Schwarz K & Cadenasso M, 2010. Mapping urban landscape heterogeneity: agreement between visual interpretation and digital classification approaches. *Journal of Landscape Ecology* 25 (1): 53–67.

Zhou W, Huang G, Troy A &Cadenasso ML 2009. Object-based land cover classification of shaded areas in high spatial resolution imagery of urban areas: a comparison study. *Remote Sensing of Environment* 113 (8): 1769–1777.

APPENDICES

The following can be found in the appendices:

Appendix A: Definitions of the features derived from the WV-2 and aerial imagery.

6.1 APPENDIX A

GLOSSARY OF FEATURE DESCRIPTIONS

The table below are the features derived from the WV-2 and aerial features for feature selection. The definitions provided are near verbatim definitions from the eCognition Developer v8.7 Reference Book.

CATERGORY	**TYPE**	**DESCRIPTION**
Layer values (page 227-229)	Mean	The mean layer intensity value of an image object
	Max difference	The difference of mean intensity of an image object
Geometry (extent) (page 251 -257)	Area	The number of pixels forming an image object
	Border length	The sum of edges of the image object shared with other image objects
	Length	The length of an image object
	Length/thickness	The length-to-thickness ratio of an image object
	Length/width	The length-to-width ratio of an image object
	Thickness	The smallest of three eigenvalues of a rectangular 3D space
	Number of pixels	The number of pixels forming an image object
	Volume	The number of voxels forming an image object
	Width	The width of an image object
Geometry (shape) (page 258 -270)	Asymmetry	The relative length of an image object, compared to a regular polygon
	Border index	It describes how jagged an image object is; the more jagged, the higher its border index
	Compactness	It describes how compact an image object is.
	Density	Describes the distribution in space of the pixels of an image object
	Main direction	The direction of the eigenvector belonging derived from the covariance matrix of the spatial distribution of the image object.

	Radius of largest enclosed ellipse	Describes how similar an image object is to an ellipse that is enlarged.
	Radius of smallest enclosed ellipse	Describes how similar an image object is to an ellipse that is scaled down.
	Shape index	Describes the smoothness of an image object border.
	Rectangular fit	Describes how well an image object fits into a cuboid of similar size and proportions.
	Roundness	Describes how similar an image object is to an ellipse by subtracting the difference of the radius of the largest enclosed ellipse from the smallest.
Texture (GLCM) (Page 329-326)	2nd Angular moment	This is another measure of local homogeneity. High values indicate high elements surrounded by smaller ones.
	Contrast	This is the opposite of homogeneity and measures local variance.
	Correlation	Measures the linear dependency of grey levels of neighbouring pixels.
	Dissimilarity	Similar to contrast, except in increases linearly and not exponentially.
	Entropy	Entropy is high when elements of the GLCM are distributed equally.
	Homogeneity	If an image is locally homogenous, values will be high.
	Mean	This is the average of the GLCM. The pixel value is not weighted by its frequency alone, but by the frequency of its occurrence in combination with a neighbouring pixel value.
	Standard deviation	Measure of dispersion around the mean.
Texture (GLDV) (Page 327 -328)	Angular second moment	This is another measure of local homogeneity.
	Contrast	Mathematically this is equivalent to GLCM Contrast
	Entropy	This is the opposite of GLDV 2nd Angular Moment.

	Mean	Mathematically this is equivalent to GLCM Dissimilarity